

Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft

Band 66

## «CaESer» - Catalog Enrichment Service

### Entwicklung eines Dienstes zur Bereitstellung von Anreicherungsdaten für Literaturdatenbanken

Heiko Jansen

Mai 2012

Fachhochschule Köln  
Fakultät für Informations- und Kommunikationswissenschaften  
Institut für Informationswissenschaft

Jansen, Heiko  
«CaESer» - Catalog Enrichment Service  
Entwicklung eines Dienstes zur Bereitstellung von Anreicherungsdaten für  
Literaturdatenbanken  
Köln: Fachhochschule Köln,  
Fakultät für Informations- und Kommunikationswissenschaften,  
Institut für Informationswissenschaft, 2012  
(Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; 66)

ISSN (elektronische Version) 1434-1115

Die Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft berichten über aktuelle Forschungsergebnisse des Instituts Informationswissenschaft der Fachhochschule Köln. Veröffentlicht werden sowohl Arbeiten der Dozentinnen und Dozenten als auch herausragende Arbeiten der Studierenden. Die Kontrolle der wissenschaftlichen Qualität der Veröffentlichungen liegt bei der Schriftleitung.  
Jeder Band erscheint in elektronischer Version (über unsere Homepage: <http://www.fbi.fh-koeln.de/institut/papers/arbeitspapiere.php>).

Fachhochschule Köln  
Fakultät für Informations- und Kommunikationswissenschaften  
Institut für Informationswissenschaft  
Claudiusstr.1 D-50678 Köln  
Tel.: 0221/8275-3376, Fax: 0221/3318583  
E-Mail: [schriftenreihe@fbi.fh-koeln.de](mailto:schriftenreihe@fbi.fh-koeln.de)

Schriftleitung: Prof. Dr. Hermann Rösch, Kerstin Wittmann  
© FH-Köln 2012

---

## INHALTSVERZEICHNIS

1	AUSGANGSLAGE .....	6
2	KATALOGANREICHERUNG IN BIBLIOTHEKEN .....	11
2.1	Begriffsdefinition .....	11
2.1.1	Katalog .....	11
2.1.2	Kataloganreicherung .....	11
2.2	Historische Entwicklung .....	13
2.3	Gegenwärtiger Stand .....	14
2.4	Technische Lösungswege.....	16
2.4.1	Integration der Anreicherungsressourcen in das Katalogisat.....	17
2.4.2	Integration der Anreicherungsressourcen in eine lokale Hilfsdatenbank.....	18
2.4.3	Integration eines externen Verweises (Hyperlink) in das Katalogisat oder eine lokale Hilfsdatenbank .....	19
2.4.4	Dynamische Integration zum Zeitpunkt der Anzeige des Katalogisats.....	20
2.4.5	Präferenzen.....	21
2.5	Situationsanalyse .....	22
3	VORBEMERKUNGEN ZU DEN WEITEREN AUSFÜHRUNGEN .....	24
4	GROB-SPEZIFIKATION.....	25
4.1	Formale und Inhaltliche Abgrenzung.....	25
4.2	Zentrale versus dezentrale Speicherung .....	26
4.2.1	Vorteile einer dezentralen Speicherung .....	26
4.2.2	Nachteile einer dezentralen Speicherung .....	27
4.2.3	Caching als Kompromiss .....	28
4.2.4	Entscheidungsprozess in der Praxis .....	28
5	INHALTE .....	30
5.1	Anzureichernde bibliographische Ressourcen .....	30
5.2	Anforderungen an Anreicherungsressourcen .....	30
5.3	Typologie der Anreicherungsressourcen.....	31
6	SAMMLUNG KOLLABORATIV ERSTELLTER DATEN.....	33
6.1	Tagging.....	33
6.1.1	Begriffsklärung.....	33
6.1.2	Zu erfassende Daten .....	34
6.1.3	Rechtliche Rahmenbedingungen.....	35
6.1.4	Verwendung erfasster Tags für die Suche.....	36
6.1.5	Verteilte Speicherung.....	37
6.2	Kommentare .....	38
6.3	Fazit .....	39
7	LIZENZEN UND NUTZUNGSRECHTE.....	41
7.1	Rahmenbedingungen .....	41

---

7.2	Begriffsklärung: Lizenzen, Verzichtserklärungen, Normen.....	42
7.3	Konsequenzen für den Datenimport.....	43
7.4	Konsequenzen für dynamisch integrierte Quellen .....	44
7.5	Nutzungsbedingungen des Anreicherungsdienstes .....	45
8	DAS PROBLEM DER IDENTITÄT.....	46
8.1	Fehlende Eindeutigkeit von Identifikatoren .....	46
8.2	Tolerante Verwendung von Identifikatoren .....	46
8.3	Bestimmung von Toleranzgrenzen.....	47
8.4	Functional Requirements for Bibliographic Records (FRBR).....	50
8.4.1	Entitätenmodell .....	50
8.4.2	Beziehungen zwischen Entitäten.....	52
8.4.3	Praktische Anwendbarkeit .....	54
8.5	Alternativen zur Gruppierung von bibliographischen Ressourcen .....	55
9	DATENHALTUNG UND DATENMODELL.....	58
9.1	Exkurs: Semantic Web, Linked (Open) Data und RDF .....	58
9.2	RDF Triple Store für die Anreicherungsdatenbank .....	61
9.3	Repräsentation der Identität von Ressourcen in RDF .....	63
9.4	Anreicherungsressourcen in RDF.....	64
9.4.1	Beispiel.....	65
9.4.2	Formale Inkonsistenzen.....	68
9.4.3	Metadaten über einzelne Anreicherungsressourcen.....	69
9.4.4	Metadaten über Teilbestände der Anreicherungsressourcen .....	69
9.4.5	Erkenntnisse und Folgerungen .....	70
9.5	Nachträglicher Ausbau der Datenbank.....	72
9.5.1	Integration von FRBR-Aussagen .....	72
9.5.2	Metadaten über die Anreicherungsressourcen .....	73
9.6	Update-Verfahren .....	73
10	INTEGRATION IN NUTZER-SYSTEME .....	75
10.1	Bereitstellung per Export und lokaler Import .....	75
10.2	Dynamische Integration .....	76
10.2.1	Workflow-Überblick .....	76
10.2.2	Widget .....	78
10.2.3	Programmierschnittstelle (API).....	79
10.2.4	Konfigurations- und Aufrufparameter für Widgets und API .....	81
10.2.5	SPARQL.....	82
11	IMPLEMENTIERUNG .....	84
12	FAZIT .....	86
13	LITERATURVERZEICHNIS .....	88
14	CODE-LISTINGS .....	100

---

## ABKÜRZUNGSVERZEICHNIS

AJA	Asynchronous JavaScript and XML
API	Application Programming Interface
CSS	Cascading Style Sheets
DOI	Digital Object Identifier
EKI	Erst-Katalogisierungs-Identifikator
FRBR	Functional Requirements for Bibliographic Records
GVK	Gemeinsamer Verbundkatalog
HTML	HyperText Markup Language
ISBN	International Standard Book Number (ISO 2108:2005)
ISRN	International Standard Report Number (ISO 10444:1994)
LCCN	Library of Congress Control Number
NBN	Nationalbibliographie-Nummer
ONIX	ONline Information eXchange
RDF	Resource Description Framework
RFC	Request for Comment
SICI	Serial Item and Contribution Identifier (ANSI/NISO Z39.56-1996)
SPARQL	SPARQL Protocol And RDF Query Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator (RFC 3986)
URN	Uniform Resource Name (RFC 1737, RFC 2141)

---

## ABBILDUNGSVERZEICHNIS

Abbildung 1 KUG Funktion «Meine Tags» nach persönlicher Anmeldung	36
Abbildung 2 Amazon: Kunden-Rezensionen zu anderer Ausgabe	48
Abbildung 3 Umschlagbilder bei LibraryThing	49
Abbildung 4 FRBR-Entitäten für bibliografische Ressourcen	51
Abbildung 5 Edition mit Identifikatoren bei Open Library	56
Abbildung 6 CaESer - Datenfluss	77

Schlagwörter: Catalog enrichment, Linked Data, Linked Open Data, Tagging, Social Tagging, Kataloganreicherung, FRBR, RDF, Dienst-Konzeption, API, Bibliographische Ressourcen, Aggregierung, Datenbank, Datenmodell, Lizenzen

---

# 1 AUSGANGSLAGE

Bibliothekskataloge – oder genauer: die WWW-Oberflächen von Bibliothekskatalogen – gelten seit längerem als nicht mehr adäquat im Bezug auf die Bedürfnisse der Benutzer. Kritisiert werden insbesondere vier Aspekte der klassischen Systeme: Layout und Präsentation, Suchfunktionalitäten, Suchgeschwindigkeit sowie Informationsumfang<sup>1</sup>.

Die folgende Arbeit beschreibt einen Ansatz zu Optimierung des Informationsumfangs. Auf die übrigen Punkte wird nur insofern eingegangen, als dass Querbeziehungen zwischen einer Umsetzung des untersuchten Ansatzes und den anderen Aspekten beleuchtet werden.

Eversberg (2002) identifiziert drei mögliche Herangehensweisen bei einer Katalogrecherche<sup>2</sup>: «Known item search», «Collocation search» und «Subject search». Eine «Known item search» ist eine Suche nach einem Dokument, zu dem bereits genaue formale Angaben bekannt sind. Unter «Collocation search» versteht Eversberg die Recherche nach alternativen jedoch logisch zugehörigen Dokumenten (z. B. andere Auflagen). Die «Subject search» schließlich soll Dokumente zu einem Thema liefern und stellt laut Eversberg die schwierigste Situation dar.

Die Aufgabe des Bibliothekskatalogs wurde zu Anfang darin gesehen, das in der jeweiligen Bibliothek vorhandene Schriftgut nachzuweisen und zugänglich zu machen (vgl. Jochum, 2007, S. 27 oder Czerwinski, 2009, S. 5 ff.). Dieser Ansatz deckt sich mit der zusätzlichen Rolle des Katalogs als internes Arbeitsinstrument zur Verwaltung und als Grundlage des Ausleihsystems, für das präzise und eindeutige Beschreibungen einzelner Medieneinheiten benötigt werden. In der Folge konzentrierte sich die Datenerfassung darauf, die Medien im Besitz der Bibliothek möglichst genau und eindeutig zu beschreiben, so dass den formalen Beschreibungsaspekten höchste Bedeutung zukam. Bibliothekskataloge eignen sich daher besonders für die «Known item search». Erst mit der Einführung von Klassifikationssystemen und darauf aufbauenden «Systematischen Katalogen» sowie später der Erfassung von Schlagwörtern in den Angaben zu den verzeichneten Medien wurden den Benutzern Mittel an die Hand gegeben, thematisch zu recherchieren und den Inhalt eines Mediums auch ohne Autopsie besser einschätzen zu können.

Der Informationsgehalt, der mit einem Klassifikationssystem ausgedrückt werden kann, ist jedoch begrenzt<sup>3</sup>. Über die Vergabe von Schlagwörtern

---

<sup>1</sup> vgl. etwa Yu und Young (2004), Mi und Weng (2008) oder Tennant (2005).

<sup>2</sup> Bzw. präziser: Bei einer Dokumenten-Recherche in Abgrenzung zu einer Fakten-Recherche.

<sup>3</sup> Mit Hilfe von komplexeren Systemen wie der Facettenklassifikation ([http://en.wikipedia.org/wiki/Colon\\_classification](http://en.wikipedia.org/wiki/Colon_classification)) oder durch den Einsatz einer detaillierten Fachklassifikation lassen sich diese Einschränkungen bis zu einem gewissen Grad überwinden. Derartige Systeme finden in Öffentlichen Bibliotheken und Hochschulbibliotheken jedoch kaum Verwendung.

---

können prinzipiell mehr Informationen transportiert werden (Eversberg, 2004a). Statistische Untersuchungen an Verbundkatalogen in Deutschland zeigen jedoch nur eine geringe Verschlagwortungsquote<sup>4</sup>. Auch scheint das Verständnis der Funktionalität von Schlagwörtern oder gar Schlagwortketten bei den typischen Benutzern von Öffentlichen Bibliotheken und Hochschulbibliotheken nicht sehr stark ausgeprägt zu sein, was sich entsprechend auf die Nutzung auswirkt (Schöllhorn, 2009, S. 39).

In der Konsequenz ist zu konstatieren, dass Bibliothekskataloge vornehmlich für die Recherche nach formalen Aspekten der verzeichneten Medien geeignet sind, den Benutzern aber wenig Mittel an die Hand geben, die für die Befriedigung des jeweiligen Informationsbedürfnisses geeigneten Medien zu finden.

Die Recherche eines Benutzers<sup>5</sup> endet jedoch insbesondere im Fall der «Subject search» nicht mit der Anzeige einer Trefferliste. Ob die gefundenen Dokumente tatsächlich für die Befriedigung des vorhandenen spezifischen Informationsbedürfnisses geeignet sind, ist nicht allein aus dem Vorkommen in der Trefferliste zu schließen. Selbst wenn die Dokumente exakt und ausschließlich das gesuchte Thema behandeln, kann die Darstellung z. B. zu detailliert oder zu grob sein. Häufig wird es jedoch bereits schwierig sein, das Interessengebiet in wenigen Suchbegriffen exakt zu beschreiben. Folglich schließt sich an die Durchführung der Recherche eine subjektive intellektuelle Evaluation der Eignung der Treffer an. Je mehr Informationen zum Inhalt der gefundenen Werke im Katalog angezeigt werden können, desto besser kann die Relevanz eingeschätzt werden. Die Studie *Online Catalogs: What Users and Librarians Want* stellt fest (OCLC Online Computer Library Center, Inc. 2009, S. 11):

Libraries need to make it easier for end users to quickly ascertain whether items meet their needs; [...] the available data needs to help users decide if it is worth their time to obtain the items [...].<sup>6</sup>

Diese Forderung gewinnt zusätzliche Bedeutung durch die Einbeziehung eines weiteren Faktors: der «Faulheit» der Benutzer. Deren zu erwartendes

---

<sup>4</sup> Seiffert (2001) ermittelt in seine Analyse der hbz-Verbunddatenbank 416.560 Vorkommen des MAB-Feldes 711 sowie 2.623.372 Vorkommen des MAB-Feldes 902. Dabei ist leider nicht klar erkennbar, ob es sich um die Anzahl der Datensätze mit Belegungen dieser Felder handelt oder um die absolute Häufigkeit der Felder. Selbst unter Annahme der ersten Variante und der zusätzlichen Annahme, dass beide Felder nie gemeinsam in einem Datensatz vorkommen, ergibt sich nur eine maximale Zahl von 3.039.932 Datensätzen und damit eine Verschlagwortungsquote von 29,92%.

<sup>5</sup> Aus Gründen der besseren Lesbarkeit wird in dieser Arbeit auf die doppelte Schreibweise für männliche und weibliche Bezeichnungen verzichtet. Alle männlichen Ausdrücke schließen auch die weibliche Form ein.

<sup>6</sup> Losgelöst von der technischen Ausgestaltung steht dieser Anspruch auch in der Tradition und entspricht dem Geist des vierten der 1931 von S. R. Ranganathan publizierten sogenannten «Fünf Gesetze der Bibliothekswissenschaft» (Ranganathan, 1989, S. 287 ff.): «Save the time of the reader / Die Zeit des Lesers sparen».



---

Verhalten wird durch das in den 1980er Jahren entwickelte «Principle of least effort» (s. Principle of least effort) beschrieben, das – angewendet auf den hier vorliegenden Kontext – in etwa besagt, dass Informationssuchende versuchen, so wenig Aufwand wie möglich in ihre Recherche zu investieren und sich mit dem ersten als minimal ausreichend angesehenen Ergebnis zufriedengeben (vgl. auch Eversberg, 2004b, Folie 1 oder Bates, 2003, S. 4 f.).

Das physische Aufsuchen einer Bibliothek stellt dabei bereits eine beträchtliche Hürde dar:

Delivery, and in fact, immediate delivery is what users have come to expect based on their experience with various downloadable media [...].

(OCLC Online Computer Library Center, Inc. 2009, S. 21).<sup>7</sup> Tatsächlich spricht bereits Coffman (1999, S. 29) im Kontext einer möglichen zukünftigen Bereitstellung digitaler Dokumente vom Anspruch auf eine «instant gratification» durch die Bibliotheksbenutzer.

Zu den Anwendungen, die mit vergleichbaren Problemen<sup>8</sup> umgehen müssen, gehören Online-Buchhändler im WWW. Zu den bekanntesten Firmen auf diesem Gebiet gehört Amazon<sup>9</sup>. Amazon verwendet mehrere Ansätze, um es seinen Kunden zu ermöglichen, sich ein möglichst detailliertes Bild von den angebotenen Medien<sup>10</sup> zu verschaffen (Emanuel, 2009, S. 119). Die Intention dahinter ist leicht erkennbar: Je sicherer ein Kunde ist, einen zu seinem Informationsbedürfnis passenden Titel gefunden zu haben, desto eher wird er bereit sein, diesen zu kaufen. Nicht von ungefähr hat die Agentur «BookNet Canada» im Jahr 2010 eine Broschüre mit dem Titel «Increase Sales & Lower Costs With Better Metadata» als Informationsschrift für Verleger veröffentlicht und einem Kapitel darin den Titel «Metadata is Marketing» gegeben (Increase Sales & Lower Costs With Better Metadata 2010, S. 4). Stimmen nach dem Kauf wiederum Erwartungshaltung und Ergebnis (Buchinhalt) überein, fördert dies die Kundenzufriedenheit und damit die Kundenbindung.

---

<sup>7</sup> Ideen zur Etablierung der Bibliothek als Ort sozialer Interaktion im Kontext von Informationsrezeption und -verarbeitung, die ein solches Aufsuchen des Ortes Bibliothek wünschenswert erscheinen lassen, sollen an dieser Stelle unbeachtet bleiben, zumal dieses Ziel kaum durch Zwänge, sondern allenfalls durch Anreize erreichbar ist.

<sup>8</sup> An die Stelle des Zeitaufwandes zum Besuch der Bibliothek tritt hier der finanzielle Aufwand zum Erwerb des Dokuments.

<sup>9</sup> <http://www.amazon.de/>.

<sup>10</sup> Die Aussagen im Text beziehen sich auf das Angebot von Medien (Bücher, CDs, DVDs). Ebenfalls bei Amazon erhältliche andere Waren sind aus der Betrachtung ausgeschlossen, obwohl die Aussagen i. d. R. unverändert übernommen werden könnten.

---

Zu den bei Amazon eingesetzten Instrumenten gehören<sup>11</sup>:

- Redaktionelle Rezensionen: Durch Amazon selbst erstellte Inhaltsbeschreibungen und Bewertungen (theoretisch neutral bzw. objektiv, in der Praxis jedoch kaum negativ).
- Verlagstexte: Inhaltsbeschreibungen und Bewertungen durch den anbietenden Verlag (i. d. R. also werbend).
- Pressespiegel: Zitiert Inhaltsbeschreibungen und Bewertungen.
- Cover-Abbildungen: Vermitteln Emotionen; stellen ggf. Beziehungen zu anderen Titeln einer Reihe her; ermöglichen ein Wiedererkennen; übermitteln eine Aussage zum Inhalt («dramatisches» Bild > spannender Krimi); enthalten oft kurze Texte mit weiteren Angaben.
- Kunden-Rezensionen: Durch Käufer erstellte Inhaltsbeschreibungen und Bewertungen; ohne Interesse, negative Aspekte zu verschweigen; oft detailliert; erhalten ihre Glaubwürdigkeit durch die jeweilige Qualität sowie durch die Anzahl vergleichbarer Bewertungen.
- Bewertungspunkte: Verdichteter Ausdruck der Kunden-Rezensionen.
- Vorschauen: Ausschnitte aus dem Buch (Inhaltsverzeichnisse, Stichwortverzeichnisse, Probekapitel, ...).
- Verwandte Titel: Diese Hinweise dienen primär der weiteren Erschließung des Angebots, sie ermöglichen aber auch eine Einordnung des vorliegenden Titels in ein Umfeld.
- Verkaufsrang: Lässt Rückschlüsse auf die Beliebtheit eines Titels zu und damit in gewissem Rahmen auf seine Qualität.
- Tags: Unkontrollierte Schlagwörter zur Beschreibung des Titels.
- Sachgebiete: Durch Amazon vorgenommene Zuordnung der Inhalte zu einem Themengebiet (Hierarchisch).

Diese Auflistung ist nicht notwendigerweise vollständig, zeigt aber deutlich, wie viel Wert Amazon auf die Bereitstellung inhaltsbeschreibender und -bewertender Informationen legt. Die formalen Metadaten fallen demgegenüber praktisch nicht ins Gewicht.

Diverse Untersuchungen belegen, dass auch die Nutzer von Bibliotheken sich inzwischen derartige Zusatzinformationen<sup>12</sup> in den für sie bereitgestellten Rechercheumgebungen wünschen – vgl. OCLC Online Computer Library Center, Inc. (2009, S. 11 ff.) oder Kneifel (2009, S. 95). Die Frage ist nun, wie Bibliotheken diesem Anspruch am besten gerecht werden können, d.h. wie eine Anreicherung von Bibliothekskatalogen um eine Vielzahl an

---

<sup>11</sup> Vgl. etwa <http://www.amazon.de/Harry-Potter-Heiligt%C3%BCmer-Todes-Band/dp/3551577773/>.

<sup>12</sup> Der Begriff «Zusatzinformation» wirkt im Hinblick auf die zuvor herausgearbeitete Bedeutung der beschreibenden Daten fast schon unangemessen. Da jedoch die formalen Metadaten für die Identifikation des passenden Datensatzes bzw. Mediums als Grundlage zwingend benötigt werden, ist er unter diesem Blickwinkel durchaus korrekt.

---

weiterführenden Informationen möglichst effizient, Ressourcen-schonend und umfassend umgesetzt werden kann.

---

## 2 KATALOGANREICHERUNG IN BIBLIOTHEKEN

### 2.1 Begriffsdefinition

Bevor die Thematik der Anreicherung von Katalogen weiter untersucht werden kann, soll zunächst der Begriff «Kataloganreicherung» oder «Catalog Enrichment» genauer definiert werden, zumal eine allgemein anerkannte feste Definition offenbar bislang nicht existiert (vgl. Samulski, 2007, S. 11 f.).

Der Begriff selbst setzt sich aus den zwei Komponenten «Katalog» und «Anreicherung» zusammen.

#### 2.1.1 Katalog

Eine ausführliche Darstellung der mit dem Begriff «Katalog» verknüpften Vorstellungen seit dem Übergang zu computergestützten Anwendungen liefert Czerwinski (2009, S. 50 ff.). Soweit in dieser Arbeit der Begriff «Katalog» verwendet wird, ist damit eine webbasierte Applikation gemeint, die insbesondere der Zielgruppe der Bibliotheksbenutzer verschiedene Mittel zur selbständigen Erschließung (formale wie thematisch-sachliche Suche) der von der Bibliothek nachgewiesenen Dokumente (egal ob in gedruckter oder elektronischer Form) an die Hand gibt. Alle weiteren Aspekte wie Informationen zum Ausleihstatus oder Standort der Literatur oder zum Konto des Benutzers spielen im vorliegenden Kontext keine oder allenfalls eine nachrangige Rolle.

Auch die Frage, wer die in dem jeweiligen System nachgewiesenen Dokumente unter welchen Aspekten und mit welchen Zielvorstellungen zusammengestellt hat, ist im Kontext dieser Arbeit unerheblich. Der zu entwickelnde Anreicherungsdienst soll keinen prinzipiellen Einschränkungen hinsichtlich der aufzunehmenden Anreicherungsressourcen unterliegen, während die nutzenden Applikation sich auf die Anzeige erwünschter Anreicherungen beschränken kann. Dies bedeutet, dass die entwickelten Ideen in ihrer Anwendung nicht auf Bibliothekskataloge beschränkt sind, wenn gleich sie dort – wie etwa aus Kapitel 1 ersichtlich – eine besondere Relevanz haben.

#### 2.1.2 Kataloganreicherung

Löhrer (2007, S. 4) schreibt unter Verweis auf Dillon und Wenzel (1989), der Begriff

[...] summiert alle zusätzlichen, nicht-bibliographischen Informationen im Bibliothekskatalog, welche Hinweise auf den Inhalt oder die thematische Abdeckung eines Buches liefern [...]. Der Begriff klammert die traditionelle Art der inhaltlichen Beschreibung durch

---

Schlagwörter oder Klassifikationen aus und umfasst neue, bisher nicht erfasste Elemente für Bücher wie Inhaltsverzeichnisse, Abstracts, Klappentexte, Umschlagbilder oder Rezensionen.

Hauer und Diedrichs (2009, S. 1033) verstehen «Kataloganreicherung» dagegen als

[...] alles was über die bisherigen Katalogregeln und „Karten“ hinausgeht und der Suche, Navigation und Information des Benutzers dient.

Sie führen dazu drei wesentliche Quellen für diese zusätzlichen Informationen auf:

1. Originaldaten, die aus dem jeweiligen Dokument / Medium extrahiert werden können,
2. maschinell oder intellektuell von Bibliothekaren oder anderen Personen aus dem jeweiligen Dokument abgeleitete Erschließungsdaten, und
3. ergänzende Daten, die sich aus nicht unmittelbar auf das einzelne Dokument bezogenen Informationen sowie aus meist von einem konkreten Dokument ausgehenden Querverweisen zusammensetzen.

Als vierten Aspekt benennen Hauer und Diedrichs (2009) zudem «Schnittstellen und Programme»:

Link-Techniken und Bookmark-Services, dann Download-Schnittstellen für lokale oder andere Literaturverwaltung und schließlich Programme zur Verarbeitung, sprich Analyse, Manipulation und Kombination von Texten, Tabellen, Graphen und Rohdaten.

Insbesondere dieser vierte Aspekt, in Teilen jedoch auch die im dritten Punkt subsumierten Funktionalitäten wie Word Clouds, gehen deutlich über das bei Löhner sichtbare Verständnis des Begriffs hinaus, das seinerseits insgesamt weiter verbreitet zu sein scheint (vgl. Samulski, 2007, S. 11 ff.). Zwar werden funktionale Erweiterungen und Verbesserungen häufig parallel zum oder unter Bezug auf den Ausbau von Inhalten diskutiert (vgl. etwa Byrum, 2005 oder Kneifel, 2009).

Dies ergibt sich nicht zuletzt daraus, dass beides einander manchmal bedingt, wie das Beispiel «Social Tagging» zeigt, bei dem eine neue Endnutzer-Funktion zur Generierung von Daten zur Inhaltserschließung dient. Jedoch dient der Begriff «Kataloganreicherung» i. d. R. nicht als Oberbegriff für den Themenkomplex. Auch erscheint dieses weite (unscharfe?) Begriffsverständnis nicht hilfreich für die Fachdiskussion, da es die präzise Benennung eines wichtigen Tätigkeitsfeldes erschwert, für das kein anderer treffender Begriff geprägt ist, während gleichzeitig Themen subsumiert

---

werden, die vornehmlich unter anderen Bezeichnungen («Web 2.0», «Library 2.0», «Soziale Software» etc.) diskutiert werden.

Im weiteren Verlauf dieser Arbeit wird das engere Verständnis des Begriffs von Löhner als Grundlage herangezogen, jedoch mit der Erweiterung, dass auch die von Bibliothekaren erstellten Erschließungselemente (Schlagwörter, Klassifikationen) als Elemente der Kataloganreicherung akzeptiert werden (zumindest im Sinne des Austausches solcher Daten zwischen getrennten Systemen). Auch die Integration von Verweisen auf direkt verfügbare Volltexte (z.B. elektronische Parallel-Veröffentlichungen, frühere Ausgaben oder – bei alten Werken mit ausgelaufenem Urrechtsschutz – alternative Versionen) oder andere Bezugsquellen wird aus Sicht dieser Arbeit als Kataloganreicherung verstanden.

## 2.2 Historische Entwicklung

Der Gedanke der Anreicherung von Bibliothekskatalogen um weitere Informationen zu den verzeichneten Dokumenten ist keineswegs neu. Und auch wenn die vorgehenden Ausführungen dies nahelegen scheinen, handelt es sich nicht um eine Idee, die erst im Kontext des World Wide Web (WWW) durch Online-Buchhändler erfunden wurde. So gab es bereits Anfang der 1980'er Jahre Untersuchungen, die zeigten, dass verschiedene Bestandteile eines Buches (Umschlag, Klappentext, erste Seite usw.) erheblichen Einfluss auf die Literaturswahl haben (vgl. Wendelin und Zinck, 1983).

Bereits Ende der siebziger / Anfang der achtziger Jahre des vergangenen Jahrhunderts wurde auch in bibliothekswissenschaftlichen Veröffentlichungen die Idee diskutiert, die Katalogisate um weitere Daten zu ergänzen, «[to] enhance the probability of users identifying the information they require» (Van Orden, 1990). Einen Überblick über die Projekte und Publikationen dieses Zeitraums bieten die Artikel von Van Orden (1990) und O'Brien (1994). Van Orden (1990, S. 27 f.) listet dabei 20 verschiedene Arten von Zusatzinformationen auf, die den Zugang zu relevanten Informationen für den Informationssuchenden sowohl erleichtern als auch erschweren könnten. Aus heutiger Sicht bzw. im Hinblick auf die aktuelle Praxis fehlen in der Liste im Wesentlichen die Umschlagabbildungen sowie die durch Benutzer vergebenen Tags oder freien Schlagwörter. Die fehlenden Umschlagabbildungen ergeben sich schlicht aus der zu jener Zeit lediglich verfügbaren Text-basierten Benutzerschnittstellen der Kataloge: die Darstellung von Bildern war technisch unmöglich. Die Idee des «Taggens» dagegen entwickelte sich erst Anfang des 21. Jahrhunderts (Tag (metadata)).

Die von Van Orden (1990, S. 31 f.) geäußerten Bedenken gegenüber der Kataloganreicherung beziehen sich im Wesentlichen auf eine Verschlechterung des Verhältnisses von Treffermenge und Genauigkeit oder – unter Verwendung der gebräuchlicheren englischen Begriffe – von «Recall» und «Precision» (vgl. dazu Precision and recall). Eine Analyse der Literatur zu

---

Evaluierungsprojekten in diesem Kontext liegt jenseits des Rahmens dieser Arbeit. In Anbetracht der Fortschritte der Suchmaschinenteknologie mit der Möglichkeit zur Indexierung großer Textmengen und der Auslieferung von Treffern in einer absteigend nach Relevanz sortierten Trefferliste, wird an dieser Stelle angenommen, dass ein Hinzufügen weiterer Daten zur indexierten Datenbasis die vom Nutzer wahrgenommene Qualität der Ergebnisse – im Sinne ihrer Eignung zur Befriedigung des vorliegenden Informationsbedürfnisses – tendenziell eher erhöht, also mehr relevante Treffer liefert bzw. die im Verhältnis zur Gesamttreffermenge relevanteren Treffer näher zum Anfang der Trefferliste verschiebt.<sup>13 14</sup> Im Hinblick auf die in Kapitel 1 beschriebene Kritik an heutigen Bibliothekskatalogen gilt diese Annahme natürlich nur insoweit, als das eine Bibliothek tatsächlich ein entsprechendes modernes Softwaresystem einsetzt.

Den Nutzen zusätzlicher Daten für die Relevanzbeurteilung der gefundenen Literatur durch den Suchenden stellt Van Orden dagegen nicht in Frage und führt hierzu auch keine möglichen schädlichen Nebenwirkungen auf. Auch Flimm (2008) sieht diesen Aspekt der «Ergebnisanreicherung» als unproblematisch und besonders sinnvoll an, während er der «Rechercheanreicherung» gegenüber noch Vorbehalte hat.

Im Laufe der Jahre wurden verschiedene Auswirkungen von Anreicherungen untersucht und es konnten eine Reihe positiver Auswirkungen festgestellt werden (vgl. Löhner, 2007, S. 10 ff.). Morris (2001, S. 34) konnte beispielsweise nachweisen, dass die Anreicherung eines Katalogisats um ein Inhaltsverzeichnis die Benutzungswahrscheinlichkeit des Titels um fast die Hälfte erhöht.

In Anbetracht der rückläufigen Erwerbungssetats von Bibliotheken (vgl. den Bericht zur Lage der Bibliotheken 2010, 2010, S. 4 f.) ist eine bessere Quote von Nutzung zu Mitteleinsatz eine besonders wünschenswerte Entwicklung.

## 2.3 Gegenwärtiger Stand

Obwohl die Kataloganreicherung somit bereits seit mehr als dreißig Jahren Gegenstand theoretischer und praktischer bibliothekswissenschaftlicher Untersuchungen ist, ist die Umsetzung des Gedankens offenbar nicht sehr weit gediehen. In den aktuellen Untersuchungen zum Thema wird jedenfalls auf eine grundsätzlich unzureichende Umsetzung verwiesen, die es zu korrigieren gilt (OCLC Online Computer Library Center, Inc. 2009, S. 53).

---

<sup>13</sup> Morgan (2011) jedenfalls sieht das Problem des «Findens» als gelöst an, und regt an, sich stärker auf die Lösung des Problems der Verarbeitung und Verwendung der Ergebnisse und der Dokumente an sich durch den Benutzer zu konzentrieren.

<sup>14</sup> Für eine grundlegende (technische) Einführung in das Thema «Information Retrieval», die Funktionsweise von Suchmaschinen und das Thema «Ranking» vgl. Manning, Raghavan und Schütze (2009).

---

Zwar gab und gibt es eine beachtliche Zahl an Projekten (vgl. etwa Samulski, 2007, S. 45 ff. oder Löhner, 2007, S. 35 ff.; eine umfangreiche Liste findet sich auch im NetbibWiki<sup>15</sup>). Jedoch scheint es dabei insgesamt an Koordination und Kooperation zu mangeln. Ein Hinweis darauf ist, dass der Datenbedarf heute existierender Kataloganreicherungen i. d. R. nicht aus von Bibliotheken angelegten Beständen gedeckt wird, sondern durch Dritte<sup>16</sup>, d. h. es ist trotz entsprechender Überlegungen (z. B. Kieft, 2004) offensichtlich nicht in größerem Maßstab zu konzertierten Datensammlungen gekommen, obwohl beispielsweise für ältere Literatur Bibliotheken die einzige Quelle von Anreicherungsdaten aus dem ursprünglichen Dokument sind (vgl. wiederum Kieft, 2004).

Die hohen Kosten für die manuelle Anreicherung waren sicher ein Grund, warum trotz entsprechenden Bewusstseins die Umsetzung der Idee nicht in größerem Maßstab realisiert wurde. Die allgemeine Weiterentwicklung der Informationstechnik (Scanner, OCR-Software usw.) und insbesondere auch die zunehmende Vernetzung von Bibliotheken mit Datenlieferanten trägt immerhin dazu bei, die aktuelle Bedeutung dieses Faktors zu minimieren. Während Byrum und Williamson (2006, S. 9) die Kosten für die manuelle Erfassung eines Inhaltsverzeichnisses mit US \$40 angeben, schlägt die Übernahme aus ONIX-Buchhandelsdaten in ihren Erhebungen lediglich mit US \$0,80 und weniger zu Buche.

Die Durchsicht der vorstehend genannten Listen zeigt auch, dass sich die Projekte im Wesentlichen im Umfeld Wissenschaftlicher Bibliotheken finden. Öffentliche Bibliotheken haben einerseits i. d. R. nicht die Ressourcen zur Durchführung von (technisch) aufwendigeren Projekten. Andererseits erlauben die im Schnitt jüngeren Bestände einen umfassenderen Rückgriff auf Daten aus dem Buchhandel.

Hinsichtlich der für Anreicherungen zu berücksichtigenden Inhalte gibt es derzeit keine allgemein akzeptierten Vorgaben oder Verbote. Große Verbreitung haben Anreicherungen mit Umschlagabbildungen (die häufig dynamisch von Amazon bezogen werden) sowie Inhaltsverzeichnissen gefunden. Eine Diskussion entspann sich hinsichtlich der Frage, ob Rezensionen als Anreicherung zulässig sind (vgl. u.a. Markner, 2005, die folgende Diskussion in der InetBib-Mailingliste<sup>17</sup> sowie Eberhardt, 2006). Ob sich daraus ein allgemein anerkannter Konsens zur Verfahrensweise innerhalb des Berufsstandes entwickelt hat, konnte nicht belegt werden. Problematisch ist eventuell auch eine einseitige Berücksichtigung bestimmter Online-Buchhändler bei der Verlinkung (ausführlicher hierzu sowie auch zur Moraldebatte Samulski, 2007, S. 70 ff.). Konkret wird Amazon aufgrund der Größe der Buch-Datenbank gerne<sup>18</sup> als Lieferant für Umschlagbilder ver-

---

<sup>15</sup> <http://wiki.netbib.de/coma/EnrichedContent>

<sup>16</sup> Zum Beispiel Amazon, Google, Syndetic Solutions oder LibraryThing.

<sup>17</sup> S. die Threads hinter <http://www.ub.uni-dortmund.de/listen/inetbib/msg27424.html> und <http://www.ub.uni-dortmund.de/listen/inetbib/msg27458.html>.

<sup>18</sup> Z. B. in den Bibliotheca2000-Systemen von OCLC und den BibDia-System von BiBer.



---

wendet, wobei angezeigten Abbildungen dann wieder Links zu Amazon sind. Der örtliche Buchhandel wird damit ebenso umgangen wie Amazon-Konkurrenten. Unklar ist, wie diese Situation besser gelöst werden könnte. Hinweise auf Buchhändler vor Ort sind bei Katalogen aufgrund der Ortsgebundenheit der Bibliotheken relativ leicht, bei institutions-unabhängigen angereicherten Literaturdatenbanken dagegen schwierig<sup>19</sup>. Eine Verlinkung zu Online-Buchhändlern dagegen bedürfte aufgrund der Menge an Anbietern wohl einer extern gepflegten Verweisdatenbank.

## 2.4 Technische Lösungswege

Für die Anreicherung von Katalogen (im Sinne von Abschnitt 2.1.1) stehen grundsätzlich vier Lösungsansätze zur Verfügung (vgl. Löhner, 2007, S. 20 ff.):

1. Integration der Anreicherungsressourcen in das Katalogisat.
2. Integration der Anreicherungsressourcen in eine lokale Hilfsdatenbank.
3. Integration eines externen Verweises (Hyperlink) in das Katalogisat.
4. Dynamische Integration zum Zeitpunkt der Anzeige des Katalogisats.

Alle Verfahren haben spezifische Vor- und Nachteile, die nachfolgend kurz erläutert werden sollen.

Der Abgleich mit potentiellen Datenquellen erfolgt in allen Szenarien i. d. R. anhand von eindeutigen Identifikatoren (z. B. ISBN), die einen effizienten Zugriff auf die Datenbank ermöglichen. Alternativ können selbstverständlich auch andere Identifikator-Varianten wie NBN, DOI, EKI usw. verwendet werden, solange sie eine bibliographische Ressource eindeutig identifizieren. Abgleiche anhand von (kombinierten) Metadaten (z. B. Autor + Titel + Jahr) sind grundsätzlich auch möglich, werden aber seltener eingesetzt, da sie a) ein erhöhtes Risiko einbringen entweder potentielle Treffer zu verpassen – z. B. anhand geringfügiger Abweichungen in der Datenerfassung (Schreibweisen, Abkürzungen) – oder falsch-positive, die Katalogbenutzer irreführende Ergebnisse zu liefern (bei hinsichtlich der formalen Daten ähnlichen Dokumenten) und b) weniger performant sind. Eine Sonderstellung nimmt hier der sogenannte «BibKey» (vgl. Voß, 2008a) ein, der einen nach formalen Regeln aus den Formalerschließungs-Metadaten synthetisierten Identifikator darstellt. Durch die Regeln zur Synthetisierung werden einige Fehlerquellen aus der Datenerfassung ausgeschlossen.

---

<sup>19</sup> Wenngleich die zunehmende Verbreitung von Geo-Lokalisierungsfunktionen anhand von IP-Adressen oder GPS-fähigen Endgeräten hier neue Möglichkeiten eröffnen könnte.

---

### 2.4.1 Integration der Anreicherungsressourcen in das Katalogisat

Die Integration von Anreicherungsressourcen direkt in das jeweilige Katalogisat ist nur für textuelle Daten möglich. Für Umschlagabbildungen, PDF-Dateien mit Abstracts usw. muss folglich ein anderer Weg gefunden werden<sup>20</sup> oder es müssen – sofern möglich – die Texte extrahiert werden, wobei i. d. R. Formatierungen verloren gehen. Je nach internen Speicherstrukturen existieren zudem Längenbeschränkungen<sup>21</sup> für die Katalogisate, so dass es bei Anreicherungsressourcen mit umfangreicheren Daten schnell zu Speicherproblemen kommt. Solange Bibliothekssysteme intern die Daten nicht in FRBR-konformen Strukturen verwalten, müssen Anreicherungen, die z. B. für mehrere separat katalogisierte Ausgaben gleichermaßen herangezogen werden können, entsprechend mehrfach abgelegt werden. Die Integration der Daten kann entweder manuell erfolgen (Copy/Paste, Abtippen), was erheblichen Personaleinsatz erfordert. Oder es erfolgt ein automatisierter Import, der jedoch – im Hinblick auf Tests der Daten und ggf. Aufbereitung des Importformates – entsprechend qualifiziertes Personal erfordert. Gerade in kleineren Öffentlichen Bibliotheken oder Spezialbibliotheken ist das notwendige Wissen oft nicht vorhanden. Soweit es nicht zu einer zentralen Aufbereitung der Anreicherungsressourcen im Rahmen eines Verbundes kommt, erfordert der Datenimport über redundanten Aufwand in jeder einzelnen Institution. Die an die deutschen Verbünde angeschlossenen (Hochschul-) Bibliotheken profitieren immerhin von einem koordinierten Scannen und Austauschen von Inhaltsverzeichnissen.<sup>22</sup>

Großer Vorteil der lokalen Speicherung ist die Nutzbarkeit der Daten für das Retrieval. Den Benutzern steht auf diese Weise ein erheblich umfangreicheres und am Inhalt der Dokumente orientiertes Einstiegsvokabular zur Verfügung. Im Fall von erfassten Inhaltsverzeichnissen werden unselbständige Werke überhaupt erst auffindbar, da diese in typischen Bibliothekskatalogen sonst nicht erfasst werden (vgl. Czerwinski, 2009, S. 22). Allerdings ist die in traditionellen Systemen verwendete Software auch nicht auf die Freitext-Suche in großen Textmengen ausgelegt, sondern meist auf eine exakte boole'sche Suche in stark strukturierten Feldern. Erst durch die Verwendung einer entsprechenden Retrieval-Komponente (z. B. einer dem eigentlichen Bibliothekssystem vorgeschalteten und von dort mit Daten versorgten Suchmaschinen-Software) können die Möglichkeiten, die die zusätzlichen Daten bieten, voll ausgeschöpft werden. Im Unterschied zur im

---

<sup>20</sup> Theoretisch wäre es möglich, beliebige binäre Datenströme z. B. per Base64-Kodierung zu serialisieren und in den Datensätzen der Katalogisate abzulegen. Dies erscheint jedoch – nicht zuletzt im Hinblick auf die Längenbeschränkungen der Datensätze – kaum praxistauglich.

<sup>21</sup> So kann ein MARC21 Datensatz maximal 99.999 Byte enthalten, die wiederum auf einzelne Datenfelder mit einer maximalen Länge von jeweils 9.999 Byte aufgeteilt werden müssen (s. MARC Standards Office, 2000).

<sup>22</sup> Vgl. <http://www.bsz-bw.de/digitalebibliothek/swbplus.html>.

---

folgenden Abschnitt 2.4.2 auf der nächsten Seite beschriebenen Vorgehensweise muss allerdings für die Verwaltung der Anreicherungsressourcen keine zusätzliche Software erworben und/oder betrieben werden.

Löhrer (2007, S. 23) führt noch wahlweise als Vor- oder Nachteil auf, dass die Anreicherungen ohne Anklicken eines weiteren Links zugänglich sind bzw. die Kataloganzeige aufblähen und für den Benutzer unübersichtlicher machen. In einem hinreichend flexiblen modernen Bibliothekssystem sollte es allerdings möglich sein, derartige Layout-Gestaltungen völlig unabhängig von der Art und dem Ort der Speicherung vorzunehmen<sup>23</sup>, so dass ein zwingender Zusammenhang zwischen der Speicherung im Katalogisat und den genannten Vor- bzw. Nachteilen nicht angenommen wird.

Als Vorteil herauszuheben ist noch, dass die Anzeige der Anreicherungen verlässlich erfolgt: es besteht keine Abhängigkeit von der Verfügbarkeit der Server eines Partners oder Dienstleisters.

Auch besteht die Option auf eine eigene Qualitätskontrolle, d.h. unbefriedigende Daten können jederzeit gezielt entfernt oder ersetzt werden. Die Kontrolle wird jedoch angesichts der zu verarbeitenden Datenmengen a priori i. d. R. nur formal erfolgen können (Test-Programme vor Einspielung der Daten). Eine inhaltliche Qualitätskontrolle wird eher a posteriori – d. h. meist durch Rückmeldungen von Benutzern – erfolgen.

Ein Beispiel stellt die Anreicherung vieler Kataloge Öffentlicher Bibliotheken um die Rezensionen der Lektoratskooperation dar. Im kommerziellen Umfeld fällt beispielsweise das Angebot «Syndetics ICE» (ICE: Indexed Content Enrichment) von Syndetic Solutions / Bowker in dieses Szenario.<sup>24</sup>

## 2.4.2 Integration der Anreicherungsressourcen in eine lokale Hilfsdatenbank

Katalogisierungssysteme sind nicht für die Aufnahme großer Datenmengen pro Katalogisat oder gar auf die Verwaltung von an ein Katalogisat angehängten Dateien ausgelegt. Eine dafür konzipierte Spezialsoftware sollte diese Funktion effizienter und ressourcenschonender erledigen, insbesondere dann, wenn sie über passende Schnittstellen zum eigentlichen Katalogisierungssystem verfügt. Im Wesentlichen sind hier die gleichen Vor- und Nachteile wie im vorhergehenden Abschnitt 2.4.1 aufzuführen, solange das Hilfssystem in der Lage ist, die Textanreicherungen an die eigentliche Retrieval-Komponente weiterzugeben.

Je nach Funktionsumfang kann eine Software für diesen Zweck weitere Vorteile bieten. Beispielsweise eine Rechtekontrolle, wenn bestimmte Zusatzdaten nur mit einer einschränkenden Lizenz bereitgestellt werden können (Campus-Lizenz eines elektronischen Volltexts zu einer Druckver-

---

<sup>23</sup> So können extern aufliegende Daten z. B. per AJAX direkt in die Anzeige des Katalogisats eingeblendet werden; vgl. auch die Abschnitte 2.4.3 und 2.4.4.

<sup>24</sup> Vgl. <http://www.bowker.com/syndetics/ice/index.html>.

---

sionen o.ä.), oder die Generierung von verkleinerten Vorschaubildern. Löhner (2007, S. 25 f.) nennt hier exemplarisch ADAM von ExLibris.

Ein Sonderfall dieses Szenarios liegt vor, wenn die Anreicherungsdaten direkt in eine dem eigentlichen Bibliothekssystem vorgelagerte Retrieval-Komponente integriert werden. Die meisten Suchmaschinen bringen Werkzeuge mit, um Texte aus einer Vielzahl von Dokumentformaten zu extrahieren. Die eigentlichen Dateien (Text-Dokumente, Grafiken usw.) müssen allerdings für die spätere Bereitstellung für die Benutzer weiterhin separat vorgehalten und verwaltet werden.

Ein Beispiel für dieses Modell stellt die Datenbank SWBplus des Bibliotheksservice-Zentrum Baden-Württemberg<sup>25</sup>, die als Quelle für Ergänzungen des Verbundkataloges betrieben wird.

#### 2.4.3 Integration eines externen Verweises (Hyperlink) in das Katalogisat oder eine lokale Hilfsdatenbank

In diesem Szenario werden nicht die Anreicherungen selbst lokal vorgehalten, sondern lediglich deren Existenz und Speicherort über einen Hyperlink im Katalogisat vermerkt. Die Komplexität des Datenimports sinkt durch die einfacheren – leichter zu verstehenden und zu prüfenden – Strukturen und reduziert sich weitgehend auf die korrekte Zuordnung der Links zum richtigen Katalogisat. Wie stark der Arbeitsaufwand im Vergleich tatsächlich sinkt, hängt von der Aufbereitung der gelieferten Daten ab: Je schlechter deren Qualität war, desto größer ist die Ersparnis.

Die Links tragen jedoch nicht zu einer Verbesserung der Suchmöglichkeiten bei. Der positive Effekt für die Benutzer liegt somit lediglich in den erweiterten Möglichkeiten zur Relevanz-Beurteilung der gefundenen Dokumente hinsichtlich des vorliegenden Informationsbedürfnisses. In den Fällen, in denen ein Lieferant von Anreicherungen die eigentlichen Daten nicht bereitstellen kann oder will, stellt dieses Vorgehen den einzigen Ausweg dar.

Besser stellt sich die Situation im Kontext einiger moderner Systeme dar, bei denen ein getrenntes Retrievalsystem (Suchmaschinen-Software) als Ergänzung zum eigentlichen Bibliothekssystem betrieben wird (beispielsweise «SISIS-Fast» oder VuFind). Diese Retrievalsysteme können bei der Übernahme der Katalogisate den enthaltenen Links folgen und die so auffindbaren Texte mit den Metadaten des Katalogisats im Suchindex zusammenführen. Bei der Verwendung dieser Funktion ist allerdings zu beachten, ob die verlinkten Daten lizenzrechtlich wirklich entsprechend verwendet werden dürfen und dass unter Umständen die Laufzeit der Indexierung rapide steigt, wenn Anreicherungsdaten von einem langsamen Server abgerufen werden müssen.

Im Vergleich zu lokal vorgehaltenen Daten ist man in diesem Szenario von der Verfügbarkeit der Dienste / Server des externen Anbieters abhän-

---

<sup>25</sup> Vgl. <http://www.bsz-bw.de/digitalebibliothek/swbplus.html>.

---

gig: stehen diese nicht zur Verfügung, können die Anreicherungen nicht abgerufen und in die Anzeige integriert werden bzw. auch nicht vom Benutzer aufgerufen werden. Eine Qualitätskontrolle der verlinkten Inhalte ist bestenfalls stichprobenartig möglich und wird ansonsten eher a posteriori durch Benutzerrückmeldungen erfolgen. Durch die Möglichkeit zur Entfernung von Links bleibt dabei eine gewisse Kontrollmöglichkeit erhalten. Da die Daten i. d. R. direkt über den Produzenten oder Aggregator abgerufen werden, sind sie zudem stets aktuell. Für die Links selbst sind allerdings regelmäßige Aktualisierungen notwendig. Dabei geht es nicht nur um die Ergänzung neuer Verlinkungen, sondern ggf. auch um das Entfernen nicht länger funktionierender Links oder um die Aktualisierung von Links, deren Struktur sich aufgrund von Änderungen in der Infrastruktur des Lieferanten geändert hat.

Da die Informationen über die verfügbaren Anreicherungen bereits bei der Aufbereitung der Ausgabe auf dem Webserver des Katalogs vorliegen, kann anders als im folgenden Szenario 2.4.4 auf die Verwendung von JavaScript im Browser des Benutzers verzichtet werden, was einen höheren Grad an Barrierefreiheit ermöglicht<sup>26</sup>.

Ein Beispiel für dieses Modell stellt die Integration der Links auf die Rezensionen von socialnet in den GVK dar<sup>27</sup>.

#### 2.4.4 Dynamische Integration zum Zeitpunkt der Anzeige des Katalogisats

In diesem Szenario geht es um die automatisierte Ermittlung von Anreicherungsressourcen zum Zeitpunkt der Anzeige der Katalogisate im Katalog.

Ein Vorteil dieser Vorgehensweise liegt darin, dass kein strukturierter Datenexport des Anbieters benötigt wird und auch keine Importe in die eigene Datenbank erfolgen müssen. Eine Optimierung der Recherche ist auf diesem Wege allerdings nicht möglich, da keine neuen Daten für den Abgleich zwischen Suchanfrage und Datenbestand hinzukommen. Umgekehrt müssen die Anbieter eine performante Schnittstelle für die Abfrage bereitstellen, was kleinere Anbieter möglicherweise überfordert oder – bei vielen Zugriffen – vor Last-Probleme stellt. Die angezeigten Daten sind stets dafür automatisch aktuell und neu vom Lieferanten bereitgestellte Daten stehen den Katalognutzern sofort zur Verfügung.

Technisch gesehen kann die Abfrage der Anreicherungsdatenbanken sowohl vor der Auslieferung der Suchergebnisse auf dem Katalogserver erfolgen als auch durch JavaScript-Code im Browser des Benutzers. Treten längere Laufzeiten bei der Abfrage der Lieferanten-Datenbanken auf, können diese allerdings die Auslieferung der Katalogergebnisse verzögern, wenn die Abfrage serverseitig erfolgt. Wird ein entsprechend kurzes Timeout für

---

<sup>26</sup> Vgl. <http://www.einfach-fuer-alle.de/artikel/bitv-reloaded/anforderung-6/bedingung-6.3/>.

<sup>27</sup> Vgl. die Kurzmeldung vom 29.11.2010 unter <http://www.gbv.de/news/neues-aus-der-verbundzentrale/2010/socialnet>.

---

die Abfragen gesetzt, können unter Umständen interessante Daten nicht berücksichtigt werden. Die dynamische Integration nicht lokal vorgehaltener Daten erfolgt daher i. d. R. über JavaScript-Code, der im Browser ausgeführt wird. Dabei können nicht nur Links zu externen Anreicherungsressourcen ermittelt und angezeigt werden, sondern diese Daten können je nach Beschaffenheit auch direkt in die Ausgabe des Katalogs integriert werden. Dies verstärkt beim Benutzer ggf. den Eindruck, dass es sich um ein Gesamtangebot des Katalogbetreibers (also der Bibliothek) handelt. Eine Qualitätskontrolle ist a priori nur stichprobenartig möglich. Durch den dynamischen Ablauf der Anreicherungen ist es zudem nur schwer möglich, einmal festgestellte unerwünschte Anreicherungen auszublenden.<sup>28</sup>

An die Stelle der Datenimporte in eine lokal betriebene Datenbank tritt in diesem Szenario die Manipulation der vom Katalog ausgelieferten Webseiten. Wie komplex diese Anpassungen sind, hängt davon ab, inwieweit der Anbieter solche lokalen Anpassungen unterstützt. Wird eine nicht-proprietäre, gut dokumentierte Template-Engine für die Generierung der Seiten verwendet und der lokale HTML-, CSS- und JavaScript-Code getrennt von den vom Hersteller ausgelieferten Bestandteilen verwaltet, gestaltet sich die Entwicklung und Pflege einfach. Andernfalls sind unter Umständen bei jedem Update der Katalogsoftware erhebliche Anstrengungen notwendig, um die lokalen Erweiterungen kompatibel zu halten und erneut zu integrieren. Vorteilhaft ist es in diesem Kontext auch, wenn Anbieter von Anreicherungen nicht nur Schnittstellen zur dynamischen Abfrage bereitstellen, sondern auch Beispiel-Code oder fertige Programme zur Nachnutzung.

Auch Bibliotheken, die an einem Verbund teilnehmen, betreiben i. d. R. einen eigenen Katalog. Dementsprechend müssen für jeden einzelnen Anreicherungsdienst die Schritte zur Integration in allen interessierten Bibliotheken individuell vollzogen werden. Ebenso müssen Aktualisierungen aufgrund von sich ändernden Schnittstellen überall nachgezogen werden.

Ein verbreitetes Beispiel für dieses Szenario ist die Anreicherung von Katalogen (z. B. OCLC Bibliotheca 2000 oder BiBer BibDia Systeme) mit Umschlagabbildungen via Amazon. Ein kommerzielles Angebot für diese Art der Anreicherung stellt das Produkt «Syndetics Plus» von Syndetic Solutions / Bowker dar.<sup>29</sup>

#### 2.4.5 Präferenzen

Die dynamische Anreicherung von Trefferanzeigen in Weboberflächen mittels JavaScript hat gegenüber den anderen Ansätzen den wesentlichen Vorteil, dass sie – vorbehaltlich der Bereitstellung ausreichender Dokumentation und des passenden JavaScript-Programmcodes – mit vergleichsweise

---

<sup>28</sup> Der JavaScript-Code müsste Sperrlisten von Identifier-/Lieferanten-Kombinationen enthalten, deren Pflege aufwendig ist.

<sup>29</sup> Vgl. <http://www.bowker.com/syndetics/plus/index.html>.

---

geringem Aufwand in die meisten am Markt befindlichen Kataloge oder Datenbanken integriert werden kann. Im Vergleich zu den Ansätzen, bei denen ein lokales Vorhalten der Objekte für die Anreicherung notwendig ist, ist zudem der laufende Pflegeaufwand zu vernachlässigen. Die Validität dieses Ansatzes wird durch die entsprechenden Geschäftsmodelle von LibraryThing und Syndetic Solutions bestätigt. Zudem entspricht dieses Vorgehen dem anhaltenden Trend, neue Anwendungen oder Optimierungen von bestehenden Anwendungen im Internet auf der Basis von «Mashups» zu entwickeln.<sup>30</sup>

In der Konsequenz sollte ein neu zu entwickelndes Angebot diesem Lösungsansatz Priorität einräumen.

Insoweit für die Bereitstellung einer entsprechenden Dienstleistung eine Sammlung von Anreicherungsressourcen erfolgt, könnten diese Daten wiederum in einem zweiten Schritt bereitgestellt werden, damit Institutionen mit ausreichenden technischen und personellen Ressourcen diese lokal zur Umsetzung eines der anderen vorstehend skizzierten Lösungswege verwenden können. Eine Bereitstellung in einem standardisierten, nicht-proprietären Format würde dabei den notwendigen Aufwand auch für diese Institutionen nachhaltig senken.

## 2.5 Situationsanalyse

Insgesamt können in der aktuellen Situation eine Reihe von aktuellen oder potentiellen Problemstellungen und Defizite identifiziert werden. Zunächst werden die verfügbaren Anreicherungsdaten nicht im größtmöglichen Umfang bzw. mit der höchstmöglichen Effizienz ausgenutzt. So werden die gescannten Inhaltsverzeichnisse zwar zwischen den Verbünden ausgetauscht, stehen aber nicht für Öffentliche Bibliotheken oder Bibliotheken mit Sammlungen deutscher Literatur im Ausland oder für Literaturdatenbanken sonstiger Produzenten zur Verfügung. Andere verfügbare Datenbestände werden unabhängig voneinander in getrennten Systemen mehrfach erfasst, wobei auch die Datenaufbereitung jeweils erneut erfolgt. Würden die Daten an einer zentralen Stelle einmal in ein homogenes Format überführt, könnten sie von dort aus erheblich leichter nach genutzt werden, insbesondere auch durch Bibliotheken außerhalb der bestehenden Verbundsysteme.

Diese zentrale homogene Datensammlung könnte einerseits als Pool für die Einspielung von Anreicherungen in lokale Datenbanken dienen. Andererseits könnte sie – eine geeignete Schnittstelle vorausgesetzt – zur dynamischen Anreicherung von Trefferanzeigen verwendet werden.

Durch die Integration auch kleiner Datenbestände z. B. von Rezensionenorganen wie IASL-Online<sup>31</sup>, rezensionen.at<sup>32</sup> oder Informationsmittel für

---

<sup>30</sup> Vgl. dazu auch Engard (2009) oder Yee (2008).

<sup>31</sup> <http://www.iaslonline.de/>.

<sup>32</sup> <http://www.rezensionen.at/>.



---

Bibliotheken (IFB)<sup>33</sup> stünden diese auch für dynamische Anreicherungen zur Verfügung, was ansonsten entweder durch fehlende Schnittstellen verhindert wird oder angesichts einer sehr geringen zu erwartenden Trefferquote kaum Sinn macht. Auch die Ergebnisse der diversen Projekte von Bibliotheken könnten so einer breiteren Nachnutzung zugeführt werden.

Die Analyse heterogener Daten und deren Aufbereitung für einen Import bindet in erheblichem Umfang qualifiziertes Personal. Ebenso sind Systembetreuer und Webmaster mit der Analyse von Schnittstellen sowie der Erstellung und Pflege von Client-Code beschäftigt. Durch die Zusammenführung verschiedener Datenbestände in ein homogenes Format und die Bereitstellung einer einzigen Schnittstelle für verschiedene Datensammlungen ergeben sich auch wirtschaftliche Vorteile für viele Bibliotheken.

Betrachtet man dieses Angebot weniger als Datenbank, denn als Dienst, so kann man den Gedanken der Zusammenführung auch in Richtung der dynamischen Anreicherung fortführen: statt nur die bereits vorab eingesammelten Daten auszuliefern, könnten auch zum Zeitpunkt der Abfrage weitere Quellen angegangen und deren Ergebnisse in eine vereinheitlichte Antwort überführt werden. Dadurch könnten Informationsbestände erschlossen werden, für die ein Gesamtabzug nicht erlaubt oder technisch nicht möglich ist. Ein Beispiel für ein vergleichbares System bildet das «RDF Book Mashup» von Christian Bizer, Richard Cyganiak und Tobias Gauß (vgl. Bizer, Cyganiak und Gauß, 2007). Nutzer der dynamischen Anreicherung profitierten damit ohne Mehraufwand von einem größeren Datenangebot.

---

<sup>33</sup> <http://ifb.bsz-bw.de/>.



---

### 3 VORBEMERKUNGEN ZU DEN WEITEREN AUSFÜHRUNGEN

Der Begriff der «Ressource» steht in den folgenden Ausführungen für ein beliebiges «Ding», das identifizierbar ist und über das Aussagen getroffen werden können. Als «bibliographische Ressourcen» werden die in Katalogen oder Literaturdatenbanken verzeichneten Ressourcen bezeichnet sowie gleichzeitig (wenn im Kontext nicht anders angegeben) die Datensätze in den Katalogen und Datenbanken, die diese Ressourcen beschreiben. Als «Anreicherungsressourcen» gelten alle Ressourcen, die der Anzeige einer Beschreibung einer bibliographischen Ressourcen beliebige weitere Informationen hinzufügen (ggf. auch als Duplikat bereits vorhandener Informationen). Der Begriff wird generell auch für die Inhalte der Datenbank des Anreicherungsdienstes verwendet, obwohl darin sowohl Anreicherungsressourcen im eigentlichen Sinne (z. B. Schlagwörter, Abstracts) als auch Informationen über Anreicherungsressourcen enthalten sind (z. B. URL von Grafikdateien). Soweit die Unterscheidung signifikant ist, sollte die Wortwahl dies widerspiegeln.

Der Begriff «Entität» wird in Abgrenzung dazu in der Regel im FRBR-Sinn (vgl. Abschnitt 8.4) verwendet, während «Objekt» im RDF-Sinn (vgl. Abschnitt 9.1) gebraucht wird.

Ausgehend von einer Skizze des grundlegenden Funktionsweise (Kapitel 4) des Dienstes werden zunächst die darin zu berücksichtigenden Inhalte beschrieben. Diese können anderswo gesammelt und im Dienst aggregiert werden (Kapitel 5) oder der Dienst selbst dient dem Aufbau solcher Sammlungen (Kapitel 6). Darauf folgt eine kurze Betrachtung der rechtlichen Aspekte der Datenaggregation (Kapitel 7).

Als Nächstes wird untersucht, wie die Zuordnung zwischen bibliographischen Ressourcen und den gesammelten Anreicherungsressourcen ablaufen (Kapitel 8). Abschließend wird untersucht, wie die gesammelten Daten organisiert und gepflegt (Kapitel 9) sowie auf welchen Wegen sie zur Anzeige gebracht werden können (Kapitel 10).

---

## 4 GROB-SPEZIFIKATION

Im weiteren Verlauf der Arbeit soll ein Entwurf für ein zentralisiertes Anreicherungssystem entwickelt werden, das bei möglichst geringem Implementierungsaufwand Lösungen für die vorstehend beschriebenen Probleme anbietet. Dazu werden technische, organisatorische und rechtliche Rahmenbedingungen analysiert und beschrieben.

### 4.1 Formale und Inhaltliche Abgrenzung

Das zu entwickelnde System soll ausschließlich zur Anreicherung von im zugreifenden (Client-) System bereits vorhandenen bibliographischen Datensätzen dienen. Es soll nicht als Fremddatenquelle für die Katalogisierung dienen und wird daher keine Daten zur Formalerschließung enthalten.

Grundsätzlich handelt es sich bei den Formalerschließungsdaten nur um Daten über ein bestimmtes Bezugswerk, die sich aus technischer Sicht nicht signifikant von einem Abstract oder Schlagwort unterscheiden. Für eine Anreicherung sind verschiedene formale Beschreibungen jedoch nicht notwendig, da sie zum einen i. d. R. in den für Endbenutzer relevanten Aspekten praktisch identisch sein sollten und zum anderen divergierende Angaben aus unterschiedlichen Quellen keinen Zusatznutzen für die Endbenutzer haben, jedoch verwirrend sein könnten. Hinzu kommt, dass im Kontext des zu entwickelnden Dienstes Informationen zu verschiedenen FRBR-Manifestationen zusammengeführt werden sollen, was potentiell eine Reihe unterschiedlicher Beschreibungen ergäbe, von denen die Mehrheit für das vom Nutzer jeweils angezeigte Werk jedoch nicht zuträfe und daher keinen Mehrwert böte. Umgekehrt allerdings könnte gerade aus dem Blickwinkel der Zusammenführung von Manifestation in Verbindung mit einer nicht-dynamischen Anreicherung (Kapitel 2.4.1 und 2.4.2) für die Erfassung bestimmter formaler Aspekte argumentiert werden, da dadurch z. B. alternative Schreibweisen des Titels (etwa bei Übersetzungen) oder des Autorennamens in den Suchindex einfließen.

Theoretisch spricht auch der häufig geringe Zusatznutzen nicht gegen die Erfassung der verschiedenen formalen Beschreibungen, da sie nicht zwingend zur Kataloganreicherung verwendet werden müssten, sondern in diesem Kontext ausgeblendet und nur in anderen Zusammenhängen berücksichtigt werden könnten. Durch die Beschränkung reduziert sich jedoch die Komplexität des internen Datenmodells zur Speicherung der gesammelten Daten und der Diskussionsbedarf zur Verwendbarkeit des Dienstes. Letztendlich handelt es sich um eine willkürliche, d.h. primär aus der Zielsetzung des zu gestaltenden Dienstes abgeleitete Entscheidung.

Das zu entwickelnde System soll zudem nicht mandantenfähig sein bzw. keinen Bezug zu seinen Client-Systemen enthalten. Mit anderen Worten soll es unabhängig von der zugreifenden oder verwendenden Stelle stets die gleichen Informationen bereitstellen. Diese Vorgabe dient ebenfalls vor-

---

nehmlich der Reduktion der Komplexität des Dienstes. Anreicherungsdaten kommerzieller Anbieter werden nicht in den Datenbestand aufgenommen, es sei denn, sie würden zentral für alle Nutzer des Anreicherungsdienstes lizenziert. Damit entfällt die Notwendigkeit, diese Daten entsprechend zu kennzeichnen, wodurch das Datenmodell vereinfacht wird. Bei der Bereitstellung / Verarbeitung von Daten für die Integration in lokale Systeme entfällt die Notwendigkeit der Ausfilterung dieser Daten im Hinblick auf die jeweiligen Berechtigungen. Es entfällt ebenfalls im Kontext der dynamischen Kataloganreicherung die Notwendigkeit, bei jedem Zugriff die Berechtigungen (i.e. die Zugehörigkeit zu einer bestimmten Institution) zu überprüfen. Angesichts des grundsätzlich beliebigen Ursprungs der Anfragen aus den Browsern der Endnutzer wäre dies durchaus nicht-trivial, da es einen komplexeren Mechanismus als einen simplen Abgleich der IP-Adresse mit einem erlaubten Adressbereich erforderte. Der meiste Aufwand wird jedoch dadurch eingespart, dass die Informationen über die Rechte nicht laufend gepflegt werden müssen.

## 4.2 Zentrale versus dezentrale Speicherung

Wie in Abschnitt 2.5 kurz angerissen, ist es nicht zwingend erforderlich, dass alle von dem zu entwerfenden Dienst bereitgestellten Daten auch vorab eingesammelt und in einer eigenen Datenbank bereitgestellt werden. Letztendlich ist dies nur eine Fortführung des bei einer dynamischen Anreicherung sowieso zugrunde liegenden Gedankens. Dementsprechend lassen sich auch viele der in den Unterabschnitten von Kapitel 2.4 diskutierten Vor- und Nachteile auf diese Fragestellung übertragen, wie die folgende Analyse zeigen wird.

### 4.2.1 Vorteile einer dezentralen Speicherung

Wesentliches Argument für die dynamische Abfrage ist die Aktualität der Daten: Änderungen in der Datenbank des Lieferanten stehen sofort allen Benutzern des Anreicherungsdienstes zur Verfügung.

Einfluss auf die Auswahl des Speicherortes könnte auch die Art und der Speicherbedarf der Anreicherungsdaten haben. Während sich Text-Daten bis zu einem gewissen Umfang problemlos in der Anreicherungsdatenbank speichern ließen, würden andere Daten wie z. B. Volltexte in PDF-Dateien oder Grafiken separat abgelegt. Die zur Schaffung einer Speicherstruktur zusätzlich erforderlichen Ressourcen – sowohl Hardware zur physischen Speicherung der Daten wie auch Software zur Verwaltung der Speicherung sowie das Personal zur Betreuung – könnten ausschlaggebend dafür sein, nur den externen Speicherort zu referenzieren und die Daten nicht im Kontext der Anreicherungsdatenbank zusätzlich vorzuhalten. Diese Daten wären grundsätzlich nicht Teil der Antwort des Anreicherungsdienstes auf Anfragen, sondern würden vom Client (Browser) sowieso in einem zweiten Schritt abgerufen, sodass der Speicherort keinen Einfluss auf die Antwort-

---

zeit des Dienstes hätte. Zugriffsbeschränkungen, Performanzprobleme oder geringes Vertrauen in generelle Verfügbarkeit des Datenlieferanten müssten dennoch bei der Entscheidung über den Speicherort und ggf. eine Duplizierung der Daten berücksichtigt werden.

#### 4.2.2 Nachteile einer dezentralen Speicherung

Welche Argumente sprechen demgegenüber für ein Vorhalten der Daten in einer eigenen Datenbank? Hier ist zunächst die Performance des Dienstes zu nennen. Die Suche in einer selbst vorgehaltenen Datenbank ist – eine einigermaßen geschickte Implementierung vorausgesetzt – schneller, als die Abfrage eines anderen Dienstes über eine Netzwerkverbindung. Potentiell kann ein nur sehr langsam antwortender Datenlieferant die Antwortzeit des gesamten Dienstes erheblich erhöhen. Technische Gegenmaßnahmen in Form von Antwortzeitbegrenzungen<sup>34</sup> der abgefragten Datenquellen sind daher zwingend erforderlich. Liegen die Antwortzeiten einer Datenquelle häufig im Bereich dieser Zeitbegrenzung, so kann die Antwort des Dienstes aufgrund wechselweiser An- und Abwesenheit der Daten dieser Quelle unzuverlässig wirken. Eine Alternative zu verhältnismäßig kurzen Timeout-Werten wäre eine Versendung der Antwort des Anreicherungsdienstes in mehreren Teilen mit Hilfe von Techniken wie «Comet»<sup>35</sup>, «WebSockets»<sup>36</sup> oder «Server-Sent Events»<sup>37</sup>. Diese Kommunikationsmethoden sind schwieriger zu implementieren als einfache Request/Response-Paare. Die Nutzung von vorhandenen Software-Bibliotheken (sofern existent) könnte den tatsächlich erforderlichen Aufwand jedoch in Grenzen halten.

Änderungen an der Abfrageschnittstelle einer Datenquelle können dazu führen, dass die von dort bezogenen Daten zeitweise komplett in der Antwort des Dienstes fehlen wenn es nicht gelingt, diese Änderungen rechtzeitig im Abfrageprogramm nachzuvollziehen. Werden die Daten vom Dienst selbst in einer Datenbank vorgehalten, fehlen dagegen bei Formatänderungen der Datenlieferung nur vorübergehend die Aktualisierungen, während der zuvor eingespielte Grundbestand weiterhin ausgeliefert werden kann. Wartungstermine der Datenquellen bleiben ebenso für den Dienst und die nutzenden Kataloge unbemerkt.

Gegen eine dynamische Abfrage / dezentrale Datenhaltung spricht auch die Tatsache, dass diese Daten nicht von der Anreicherungsdatenbank für eine Indexierung in lokale Kataloge weitergegeben werden kann und daher nur für die dynamische Anreicherung von Katalogen zur Verfügung stehen.

---

<sup>34</sup> Auch als «Timeout» bezeichnet; Angabe, wie lange maximal auf eine Antwort der Gegenseite gewartet wird.

<sup>35</sup> Vgl. [http://en.wikipedia.org/wiki/Comet\\_%28programming%29](http://en.wikipedia.org/wiki/Comet_%28programming%29).

<sup>36</sup> Vgl. <http://en.wikipedia.org/wiki/WebSockets>.

<sup>37</sup> Vgl. <http://dev.w3.org/html5/eventsources/>.

---

### 4.2.3 Caching als Kompromiss

Eine Methode, um Probleme in der Kommunikation mit Datenlieferanten zu reduzieren, wäre das Zwischenspeichern («cachen») von Daten: Einmal über dynamische Anfragen abgeholte Daten werden für einen definierbaren Zeitraum im Dienst vorgehalten. Weitere Anfragen an den Dienst für die gleichen Daten werden dann während des genannten Zeitraums direkt mit Hilfe der Daten aus dem Zwischenspeicher beantwortet. Erst nach Ablauf der Frist werden die Daten wieder vom ursprünglichen Lieferanten geholt und erneut zwischengespeichert. Erst mit dieser verzögerten Anfrage bei eigentlichen Lieferanten gelangen dann allerdings Aktualisierungen bis zum Dienst und damit zu den angereicherten Katalogen. Ein Vorteil dieses Zwischenspeichers liegt auch in der Reduktion der Anfragelast, die beim Datenlieferanten ankommt. Andererseits erhöht die Integration eines Zwischenspeichers die Komplexität der Dienst-Software, wenngleich der notwendige Aufwand relativ gering sein dürfte, da es sich um eine häufig auftretende Anforderung handelt für deren Bewältigung generische Software existiert.

Allerdings müssen in jedem Fall die Lizenzbedingungen des jeweiligen Lieferanten daraufhin geprüft werden, ob erstens eine indirekte Abfrage – mit dem in dieser Arbeit beschriebenen Anreicherungsdienst als Mittler zwischen Client (Browser) und Lieferant – überhaupt erlaubt ist und zweitens ob solcherart vermittelte Daten zwischengespeichert werden dürfen. LibraryThing beispielsweise untersagt beides für einen Teil seiner Abfrageschnittstellen.<sup>38</sup>

### 4.2.4 Entscheidungsprozess in der Praxis

In der Praxis wird eine Reihe von Faktoren die Entscheidung beeinflussen, sofern die Rahmenbedingungen überhaupt eine Wahl lassen. Steht keine bedarfsgerechte API oder kein Datenexport zur Verfügung, ist das Vorgehen ebenso festgelegt wie in jenen Fällen, in denen lizenzrechtliche Bestimmungen oder Benutzungsbedingungen eine prinzipiell mögliche Implementierung entsprechend reglementieren. Dies muss keine absolute Beschränkung sein: Ist für eine API nur eine bestimmte Menge an Zugriffen pro Zeiteinheit erlaubt und unterschreitet diese Menge den angenommenen Bedarf, so ist der Weg über die dynamische Abfrage effektiv versperrt.

Ein «weicher» Auswahlfaktor ist die vom Datenlieferanten bereitgestellte Datenmenge. Ist diese sehr groß, macht die dynamische Anreicherung unter Umständen Sinn, weil die für den Anreicherungsdienst verfügbare Hard- und Software überfordert sein könnte. Ist sie sehr klein, so spricht dies für eine Übernahme in die eigene Datenbank, da das Verhältnis von Treffern zu Anfragen an den Lieferanten sehr ungünstig ausfallen wird und somit unnütz Last auf beiden Seiten erzeugt wird.

---

<sup>38</sup> Vgl. [http://www.librarything.com/wiki/index.php/LibraryThing\\_APIs](http://www.librarything.com/wiki/index.php/LibraryThing_APIs).

---

Ein weiterer Auswahlfaktor ist das Antwortzeitverhalten des Datenlieferanten: Eine schlechte Performance spricht für die Übernahme in die zentrale Anreicherungsdatenbank. Schwer zu formalisieren, aber ebenfalls wichtig ist die allgemeine Einschätzung des Lieferanten in Bezug auf eine unterbrechungsfreie Verfügbarkeit und dauerhafte Existenz. Bestehen hier Zweifel, sollten die Daten in die zentrale Anreicherungsdatenbank übernommen werden. Schließlich wäre für die Übernahme in die zentrale Anreicherungsdatenbank noch zu prüfen, wie weit der Prozess der Datenübernahme im Hinblick auf Aktualisierungen automatisierbar ist. Ist hierzu regelmäßig persönliche Interaktion und manuelles Anpassen der Auswertung und Übernahme der Daten notwendig, so stellt der Zugriff auf eine stabile API eine geeignetere Alternative dar.

Eine eindeutige, formalisierbare Handlungsanweisung kann letztlich nicht a priori formuliert werden, da in der Praxis mehrere der genannten Faktoren gegeneinander abzuwägen sind. Die Entscheidung wird notwendigerweise von der Erfahrung und der darauf beruhenden subjektiven Situationseinschätzung des Entscheidungsträgers abhängen.

---

## 5 INHALTE

Für die technische Entwicklung des Anreicherungsdienstes wäre es hilfreich zu wissen, was für Daten verwaltet werden müssen. Nachfolgend wird untersucht, welche Informationen darüber aus den Rahmenbedingungen, unter denen der Dienst verwendet werden soll, sowie aus bestehenden Standards und Publikationen abgeleitet werden können.

### 5.1 Anzureichernde bibliographische Ressourcen

Welche Daten von dem Anreicherungsdienst letztlich bereitgestellt oder referenziert werden, ist abhängig von den bibliographischen Ressourcen, zu denen weitere Daten geliefert werden sollen. Im engeren Sinne einer «Kataloganreicherung» sind dies primär bibliographisch selbständige Dokumente wie (gedruckte, elektronische und Audio-) Bücher oder Musik-Tonträger und Film-DVD/CD. Bei der Anreicherung einer allgemeinen Literaturdatenbank können es aber ebenso gut die Nachweise von Beiträgen aus Sammelwerken oder Zeitschriftenartikel sein (vgl. die Projektidee und Untersuchung von Page, 2010). Im Zentrum der folgenden Überlegungen wird aus praktischen Gründen – dies meint die Anzahl der in diesem Kontext existierenden Projekte und Erfahrungen – die Anreicherung von typischerweise in Bibliothekskatalogen verzeichneten Dokumenten stehen.

Wie im Kapitel 2.4 gezeigt wird, ist es für die Anreicherung wichtig, dass (möglichst eindeutige) Identifikatoren der anzureichernden bibliographischen Ressourcen vorliegen. Dies ist bei einer Vielzahl von Dokumenttypen der Fall<sup>39</sup>, sodass eine grundsätzliche Beschränkung auf Anreicherungen zu bestimmten Dokumenttypen unter diesem Gesichtspunkt unnötig erscheint.

### 5.2 Anforderungen an Anreicherungsressourcen

Der grundlegende Mechanismus des Anreicherungsdienstes besteht darin, auf eine Anfrage hin entweder Informationen textueller Art direkt zurückzugeben oder alternativ Verweise (URLs) auf den Speicherort weiterer Informationen zu liefern. Letztere können dann zu einem beliebigen späteren Zeitpunkt in einem getrennten, unabhängigen Prozess abgerufen werden. Die einzige Anforderung an durch den Dienst zu vermittelnde Anreicherungsressourcen besteht folglich darin, in irgendeiner Weise als digitales Objekt repräsentierbar zu sein. Unter diesen Umständen ist nicht damit zu rechnen, dass schon zum Zeitpunkt der Konzeption des Dienstes alle später zu verzeichnenden Arten von Ressourcen festlegen werden können. Das Datenmodell zur Speicherung muss daher flexibel erweiterbar sein.

---

<sup>39</sup> Zum Beispiel ISBN, EKI, LCCN usw. bei Büchern; DOI oder SICI bei Artikeln; ISRN bei Reports.

---

Allerdings müssen für die auszuliefernden Daten Typ-Bezeichner festgelegt und dokumentiert werden, damit die Verwender des Dienstes entscheiden können, welche Daten sie zur Anzeige bringen und wie diese Anzeige erfolgen soll. Zum Beispiel könnte ein Anwender auf die Anzeige von Rezensionen oder auf fremdsprachiges Material verzichten wollen. Und ein Link auf ein extern aufliegendes Abstract muss als solcher per HTML-Markup ausgezeichnet werden, damit er direkt aufrufbar ist, während ein mitgelieferter Abstract-Text direkt zur Anzeige kommen kann. Es ist daher sinnvoll, die eigentlichen Anreicherungsdaten um zusätzliche Metadaten über die jeweilige Anreicherungsressource zu ergänzen. Für die einzelnen Elemente können dabei teilweise unterschiedliche Metadaten (Dateiformat, Sprache, Ersteller usw.) relevant sein, was die Komplexität und die Ansprüche an die Flexibilität des Datenmodells zusätzlich erhöht.

### 5.3 Typologie der Anreicherungsressourcen

Als Ausgangsbasis bietet es sich an, auf existierende Zusammenstellungen von möglichen Anreicherungsressourcen zurückzugreifen. Eine möglichst breite Basis erhöht die Wahrscheinlichkeit, Sonderanforderungen bestimmter Typen an Metadaten für ihre Verarbeitung / Verwendung frühzeitig erkennen und im abstrakten Datenmodell berücksichtigen zu können. In Frage kommen hier zum Beispiel die Liste bei Van Orden (1990, S. 28), die Zusammenstellung durch die Cataloging Enrichment Initiative (2004) und insbesondere die Elemente im ONIX-Datenformat (ONIX for Books, 2009). Tatsächlich lehnt sich auch die Zusammenstellung möglicher Anreicherungsressourcen durch eine Arbeitsgruppe der Arbeitsgemeinschaft der Verbundsysteme an eine entsprechende Liste aus dem ONIX-Standard an (vgl. Arbeitsgruppe Kooperative Verbundanwendungen der Arbeitsgemeinschaft der Verbundsysteme (AG KVA), 2010)<sup>40</sup>. Das ONIX-Format insgesamt geht aber über diese Auswahl hinaus und enthält Felder für eine ganze Reihe weiterer Angaben, die im Rahmen einer Kataloganreicherung von Interesse sind. Beispielsweise sind Angaben zur intendierten Zielgruppe (Elementgruppe P.13, vgl. ONIX for Books, 2009, S. 71 ff.) oder über Preise, für die der jeweilige Titel nominiert war bzw. die er gewonnen hat (Elementgruppe P.17, vgl. ONIX for Books, 2009, S. 86 ff.) möglich.

LibraryThing zeigt weitere Optionen auf, wenn es beispielsweise die Angabe von wichtigen Orten oder von Charakteren, die mit dem Inhalt einer bibliographischen Ressource in Verbindung stehen, erlaubt.<sup>41</sup>

---

<sup>40</sup> Die AGKVA stellt hier eine Beziehung zur «ONIX-Liste 33 der ONIX-Version 2.1» her, die hinsichtlich ihrer Formulierung eine Verknüpfung der ONIX-Version und des Inhaltes der Code-Liste nahe legt. Ein solcher Zusammenhang existiert jedoch nicht (mehr), denn die Code-Listen werden unabhängig von der Version des Standards gepflegt: «From Release 2.1 of ONIX for Books, the ONIX Code Lists are being maintained separately from any of the individual ONIX message formats» (vgl. <http://www.editeur.org/14/Code-Lists/>).

<sup>41</sup> Vgl. etwa <http://www.librarything.com/work/3093889/commonknowledge>.



---

Abseits klassischer Anreicherungsressourcen beklagen Hauke u.a. die mangelnde und inkonsistente Berücksichtigung von elektronischen Parallelausgaben in Bibliothekskatalogen (vgl. Hauke, Hötzeltdt und Rumler, 2010 bzw. Hauke, Rumler und Hötzeltdt, 2010). Auch die Lösung dieses Problems durch die Ergänzung entsprechender Verweise kann als Aufgabe einer Kataloganreicherung betrachtet werden.

In Anbetracht der im Kapitel 1 beschriebenen aktuellen Überlegenheit zumindest bestimmter Online-Buchhändler im Hinblick auf die Versorgung ihrer Kunden mit ergänzenden Informationen erscheint es insgesamt sinnvoll, sich für den Aufbau eines Anreicherungsdienstes an den Standards dieses Dienstleistungssektors zu orientieren.

Geo-Koordinaten mit Bezug auf das Ausgangsdokument oder gar darin behandelte Gen-Sequenzen, die Page (2010) als zusätzliche Metadaten vorschlägt, finden sich in diesen Listen allerdings nicht. Gerade bei hochspezialisierten Fachartikeln besteht die Möglichkeit, dass die Fachöffentlichkeit des jeweiligen Forschungsgebiets Ideen zur Anreicherung der bibliographischen Beschreibung hat, auf die Bibliothekare nicht kämen. Daraus folgt, dass unabhängig vom Ausgangspunkt das dem Datenspeicher zugrundeliegende Datenmodell auf eine flexible Erweiterbarkeit hin ausgelegt werden muss.

Während die Berücksichtigung zum Beispiel von Literaturpreisen und (Alters-) Zielgruppen eher in Richtung Belletristik und der Erfüllung der Bedürfnisse Öffentlicher Bibliotheken zielt, böte eine fachwissenschaftlich ausgerichtete Anreicherung Mehrwerte insbesondere für Hochschul- und Spezialbibliotheken.

---

## 6 SAMMLUNG KOLLABORATIV ERSTELLTER DATEN

Der bislang beschriebene Dienst ist dafür gedacht, einer beliebigen Anzahl von Institutionen als Datenquelle zu dienen, indem er – a priori und vollständig oder dynamisch und auszugsweise – bestehende Datensammlungen zusammenführt. Es spricht jedoch nichts dagegen, die Datenbank des Dienstes auch als Primärdatenbank zu betrachten und zu verwenden, d. h. Anreicherungsressourcen direkt in ihr zu sammeln, die nicht bereits zuvor an einer anderen Stelle aggregiert wurden. Ein konkretes, nachstehend näher erläutertes Anwendungsszenario besteht in der Verwendung der Datenbank als Sammelstelle und Austauschinstrument für Tags oder Kommentare. Nachstehend werden Möglichkeiten, Probleme, Einschränkungen und alternative Ansätze für diese beiden Nutzungsszenarien skizziert.

Dabei ist zu beachten, dass Tags und Kommentare nicht die einzigen Daten sind, die aus einem kollaborativen Prozess gesammelt werden können. Beispielsweise ist ein Bewertungssystem mit der Vergabe von Punkten auf einer Bewertungsskala wie bei Amazon oder eine schlichtere Positiv-Bewertung analog der «Like»-Funktion von Facebook oder Googles «+1» vorstellbar. Die folgenden Ausführungen können daher nicht als abschließende Diskussion aller Optionen oder Schwierigkeiten betrachtet werden, sollen aber die grundsätzlichen Fragen beleuchten.

### 6.1 Tagging

#### 6.1.1 Begriffsklärung

Unter «Tagging» verstehen wir die Vergabe von Termen oder Bezeichnern durch beliebige Personen (in Abgrenzung zu professionell mit dieser Aufgabe betreuten Personen) an beliebige Ressourcen (Bilder, Webseiten, Bücher o.ä.) im World Wide Web (vgl. Thomas, Caudle und Schmitz, 2010, S. 223, Kneifel, 2009, S. 32 ff. und Tag (metadata)). Der Prozess selbst wird auch als «collaborative tagging», «social classification», «social indexing» oder «social tagging» bezeichnet (vgl. Folksonomy), wenn die vergebenen Tags auch für andere Personen sicht- und nutzbar sind. Dies wird im Kontext der vorliegenden Arbeit vorausgesetzt, so dass die genannten Begriffe hier tatsächlich als synonym gelten können. Häufig müssen die Tags aus einem einzigen Wort bzw. genauer: aus einer Leerzeichen-freien Kette von Zeichen (Buchstaben, Zahlen) bestehen, wobei zur Übernahme komplexerer Begriffe durch die taggenden Personen gelegentlich Leerzeichen durch Unterstriche ersetzt werden. Ziel des «taggens» ist die Kategorisierung der Ressourcen und ihre Kennzeichnung mit dem der jeweiligen Person eigenen

---

Vokabular<sup>42</sup> zum leichteren späteren Wiederauffinden und Einordnen (vgl. Thomas, Caudle und Schmitz, 2010, S. 223 ff.; Tag (metadata)).

### 6.1.2 Zu erfassende Daten

Die einfachste Annahme besteht darin, bei jedem Tagging-Ereignis einen Datensatz in die Anreicherungsdatenbank hinzuzufügen, der die Zeichenkette des Tags mit dem/den Identifikator(en) des getaggtten Dokuments bzw. deren Aggregation in Verbindung setzt. Diese Vorgehensweise ist valide, würde jedoch zu einer Reihe von Einschränkungen in der späteren Nutzung der gesammelten Tags führen.

Im ersten Schritt wäre zu überlegen, ob die Tags bei der Aufnahme in die Datenbank «normalisiert» werden sollten (vgl. z. B. Thomas, Caudle und Schmitz, 2010, S. 224 f. oder Kneifel, 2009, S. 32 ff.). Ein sehr einfacher Fall ist die Änderung von Groß- in Kleinschreibung. Die «Normalisierung» könnte mittels linguistischer Verfahren oder anhand von Wörterbüchern z. B. auch dazu führen, dass immer nur eine Singular-Form zugelassen wird, dass Komposita zerlegt oder erzeugt werden oder dass rein persönliche Tags («gelesen») abgelehnt werden<sup>43</sup>.

Ebenso ist eine Rechtschreibkontrolle denkbar. Zentraler Zweck der Normalisierung wäre eine Erhöhung des Recalls bei der späteren Verwendung der Terme in der Suche, indem Schreibvarianten des gleichen Begriffs auf eine Vorzugsform abgebildet werden. Eine solche Manipulation der Benutzereingabe müsste diesem allerdings unmittelbar angezeigt werden, um Irritationen bei der weiteren Nutzung zu vermeiden. Da die Tags – wie in Abschnitt 6.1.1 beschrieben – aus dem Vokabular des Benutzers stammen, sind invasivere Mechanismen wie eine Synonym-Zusammenführung auf einen Vorzugsbegriff unangebracht, da damit die Zugänglichkeit für die Benutzer erheblich eingeschränkt würde. Auf der Basis eines Grundbestands an Tags oder eines Wörterbuches wäre ein Vorschlagsassistent bei der Eingabe von neuen Tags eine geschickte Möglichkeit, auch ohne eine nachträgliche Normalisierung die Menge an lediglich geringfügig voneinander abweichenden Tags zu reduzieren.

Anstelle des Anlegens eines neuen Datensatzes für jeden Tagging-Vorgang könnte für jede bereits im System vorhandene Zeichenkette ein Zähler inkrementiert werden. Im RDF-Kontext wäre es dazu notwendig, wenn jede neue Zeichenkette eine URI erhielte, um den Zähler zuordnen zu können. Durch die Verwendung eines Zählers statt separater Datensätze würde die Anzahl der Datensätze in der Datenbank in Abhängigkeit von der Anzahl der Duplikate gesenkt. Hier hätte die Normalisierung also einen zusätzlichen positiven Effekt im Rahmen der Datenhaltung. Erfolgt die Prä-

---

<sup>42</sup> Dies stellt eine wesentliche Abgrenzung zur Verschlagwortung mit einem kontrollierten Vokabular dar (Mendes, Quinonez-Skinner und Skaggs, 2009, S. 32).

<sup>43</sup> Eine solche Ausfilterung persönlicher Tags erfolgt z. B. bei LibraryThing (Mendes, Quinonez-Skinner und Skaggs, 2009, S. 33).

---

gung der URI allein auf Basis der Zeichenkette geht in diesem Prozess allerdings die Information verloren, wie oft ein bestimmtes Tag für ein bestimmtes Dokument oder Werk vergeben wurde. Um diese Information zu erhalten, müsste die URI über die Kombination von Tag und Identifikator vergeben werden. In diesem Fall müsste ein zusätzlicher Mechanismus für die Ermittlung der Gesamt-Häufigkeit implementiert werden.

Weitere Metadaten zum Tagging, deren Erfassung unter bestimmten Gesichtspunkten sinnvoll wäre, die aber vergleichbare Probleme wie Erfassung der Vergabehäufigkeit mit sich brächten, sind die Angaben zum Ursprungssystem<sup>44</sup>, zum Erstellungsdatum oder gar zur Benutzerkennung der taggenden Person. Nur durch die Erfassung der Benutzerkennung ist beispielsweise eine Funktion wie «Zeige mir die von mir vergebenen Tags» realisierbar, wie sie etwa vom Kölner UniversitätsGesamtkatalog (KUG)<sup>45</sup> angeboten wird.

### 6.1.3 Rechtliche Rahmenbedingungen

Die Erfassung von personalisierten Daten führt zu veränderten Rahmenbedingungen.

Spiteri (2007, S. 23) und Kneifel (2009, S. 34, 36 und 134) weisen grundsätzlich darauf hin, dass eine Bibliothek, die in ihrer Anwendung Tagging zulässt, eine Benutzungsrichtlinie haben sollte, die es erlaubt, problematische (strafrechtliche bedenkliche, obszöne oder auch einfach nur sinnfreie) Tags zu löschen. Entsprechendes gilt auch bei der Verwendung des Anreicherungsdienstes als Speicherort. Eventuell müsste auch schon die asynchrone spätere Weitergabe von zunächst lokal erfassten Tags an den Anreicherungsdienst durch die Zustimmung der Benutzer zu einer entsprechenden Benutzungsordnung abgedeckt werden.

---

<sup>44</sup> In welchem Katalog / welcher Datenbank wurde das Tag vergeben?

<sup>45</sup> <http://kug.ub.uni-koeln.de/>.



**Abbildung 1 KUG Funktion «Meine Tags» nach persönlicher Anmeldung**

Unklar ist auch, ob die Anwender gleichzeitig zusätzliche separate Benutzungsbedingungen des Anreicherungsdienstes akzeptieren müssten.

Die Weitergabe von Benutzerkennungen zur Verknüpfung von Tags (oder anderen Daten – vgl. den folgenden Abschnitt 6.2) mit ihren Erstellern aus dem unmittelbaren Kontrollbereich der jeweiligen Betreiber-Institution (Bibliothek) an einen externen Dienstleister geht jedoch noch einen Schritt weiter.

Die Klärung der Frage, unter welchen rechtlichen Rahmenbedingungen dieses Vorhaben umsetzbar wäre, sprengt allerdings den Rahmen der vorliegenden Arbeit und muss separat erfolgen.

#### 6.1.4 Verwendung erfasster Tags für die Suche

Tags werden primär für zwei Zwecke verwendet. Zum einen dienen sie stets dazu, ein Dokument in irgendeiner Weise näher zu beschreiben, indem ein neues Datum (eine Zeichenkette) zu der zum Zeitpunkt der Tag-Zuordnung vorhandenen Beschreibung hinzugefügt und anschließend bei jedem Datensatzabruf mit angezeigt wird. Zum anderen sind diese Zeichenketten in den meisten Fällen anschließend suchbar, d. h. es ist möglich über die Suche nach dem vergebenen Tag alle mit diesem Tag versehenen Dokumente aufzufinden.

Die Suche ist technisch problemlos realisierbar wenn die Tags direkt in die lokale Datenbank integriert werden, da ein Verweis von einem Tag zu einem oder mehreren Datensätzen mit den Mitteln jeglicher aktueller Datenbanktechnologie einfach abbildbar ist.

Im technisch einfachsten Szenario unter Beteiligung des Anreicherungsdienstes würden die Tags über die lokale Anwendung (Web-Katalog o.ä.) erfasst, jedoch nur im Anreicherungsdienst gespeichert und von dort dynamisch für die Anzeige in den getaggten Datensätzen abgerufen. Eine Suche nach den mit einem bestimmten Tag versehenen Datensätzen erforderte unter diesen Umständen jedoch einen komplexen Ablauf. Zunächst müsste der

---

Anreicherungsdienst eine Liste aller Identifikatoren (ggf. gefiltert nach jenen Typen, die in der jeweiligen lokalen Datenbank vorhanden sind) zurückliefern, denen das Tag zugeordnet wurde. Im zweiten Schritt müsste eine Suche mit einer – theoretisch beliebig umfangreichen – ODER-Verknüpfung der gelieferten Identifikatoren über den lokalen Datenbestand ausgeführt werden. Die übrigen Abfragen des Anreicherungsdienstes sind darauf ausgelegt, von Identifikatoren ausgehend Daten zu ermitteln. Im Fall der Tag-Suche ist die Fragerichtung umgekehrt, was ggf. eine Performance-Optimierung erschweren könnte. Im Hinblick auf die lokale Datenbank ist nicht auszuschließen, dass eine Abfrage über hunderte oder tausende Identifikatoren (bei häufig vergebenen Tags) massive Probleme verursachen oder zumindest erhebliche Zeit in Anspruch nehmen könnte.

Daraus folgt, dass zur Umsetzung einer Suche über Tags die von den Nutzern einer Datenbank vergebenen Tags nicht nur zur Speicherung an den Anreicherungsdienst übergeben, sondern in jedem Fall auch lokal indexiert werden sollten. Zudem sollten die zu lokal vorhandenen Dokumenten im Anreicherungsdienst gesammelten Tags über regelmäßige Importe entsprechender Auszüge der Anreicherungsdatenbank dem lokalen Datenbestand hinzugefügt werden.

Einen anderen Weg geht LibraryThing bei «LibraryThing for Libraries» (LTFL). Die Anzeige der Tags erfolgt per JavaScript-basierter, dynamischer Abfrage vom LibraryThing-Server. Es werden jedoch nur die für den eigenen Bestand passenden Tags angezeigt indem die in Frage kommenden Tags auf dem LibraryThing-Server gegen die Bestände des jeweiligen Kunden (der Bibliothek) abgeprüft werden. Damit das möglich ist, muss die Bibliothek zuvor die ISBN aus dem eigenen Katalog extrahieren und auf den LTFL-Server hochladen (vgl. Mendes, Quinonez-Skinner und Skaggs, 2009, S. 34). Eine lokale Suche über die Tags ist nicht umsetzbar, weil ein vollständiges Herunterladen dieser für das Geschäftsmodell von LibraryThing essentiellen Daten nicht möglich ist.

### 6.1.5 Verteilte Speicherung

Neben der Suche können mit einer Tagging-Komponente – wie vorstehend in Abschnitt 6.1.2 beschrieben – weitere Funktionen für die Benutzer verknüpft werden. Beispiele sind: «Zeige mir eine Liste meiner Tags» oder «Zeige nur Dokumente, denen ich selbst dieses Tag zugeordnet habe». Analog zur lokalen Indexierung der Tags für die Zwecke der Suche könnte sich ein Ausweg aus der Datenschutzproblematik und eine Vereinfachung des Datenmodells in der zentralen Anreicherungsdatenbank daraus ergeben, dass die zu erfassenden Daten verteilt gespeichert werden bzw. nur Teile (die reinen Tags) in die zentrale Datenbank repliziert werden. Mit anderen Worten würden zunächst alle gewünschten, bei der Vergabe eines Tags anfallenden Informationen in der lokalen Datenbank erfasst und erst nachträglich lediglich das Tag in den Anreicherungsdienst übertragen.

---

## 6.2 Kommentare

Greenall (2011) beschreibt ein Modell für das kollaborative Kommentieren von Dokumenten / Werken. Er greift für die Implementierung auf einen existierenden, weit verbreiteten Dienstleister zurück, der üblicherweise für die Kommentarfunktion in Blogs o.ä. verwendet wird. Die Hürde zur Verwendung dieses Dienstes lag sehr niedrig, da er keinerlei lokal zu installierende Komponente benötigt, sondern lediglich über kleine Anpassungen in den Oberflächen-Templates und die Integration einiger JavaScript-Bibliotheken in bestehende Web-Anwendungen integriert werden kann. Die Bereitstellung der gesamten Kommentarfunktion ist nicht zuletzt deshalb von großer Bedeutung, weil diese nochmals erheblich aufwendiger ist als die Tagging-Funktion. So können Kommentare untereinander in Beziehung stehen (Kommentare können Antworten auf frühere Kommentare sein; mehrfache Antworten können sich zu Diskussionen entwickeln). Wie Tags mit der Anzahl ihrer Verwendungen weisen Kommentare auch eigene Metadaten auf. Beispielsweise erlaubt es der von Greenall verwendete Dienst, einzelne Kommentare zu bewerten.

Analog zu dem in der vorliegenden Arbeit beschriebenen Anreicherungsdienst würde die von Greenall entwickelte Kommentierungs-Lösung von Identifikatoren für Werke profitieren. Als beste verfügbare Näherung für so ein System verwendet Greenall die «OCLC Work Number (OCI)», die zuvor anhand einer vorhandenen ISBN ermittelt wird. Die mit der Verwendung von ISBN verbundenen Einschränkungen sind bekannt, werden aber zugunsten der praktischen Umsetzbarkeit in Kauf genommen (Greenall, 2011, Abschnitt "Work-ID System").

Auch bei der Speicherung der Kommentare könnte eine Aufteilung der Daten in personenbezogene und sonstige Informationen erfolgen, wobei die personenbezogenen Daten das lokale System, in dem die Kommentierung erfolgte, nicht verlassen würden. Bei Greenalls Verwendung eines unabhängigen Systems dagegen gibt es keine lokalen Daten; es ist zwar möglich, einen Kommentar unter einer bestimmten Benutzerkennung abzugeben, aber diese ist Teil des Kommentar-Dienstes.

Ein möglicher Nachteil an der von Greenall erdachten Lösung ist, dass die Kommentare nicht für eine Volltextsuche in den lokalen Systemen herangezogen werden können.

Bemerkenswert dagegen ist, dass sich die Kommentare und die damit verbundenen Informationen über die Kommentierenden automatisch über eine Vielzahl von anderen Anwendersystemen verbreiten. Zwar wird es sich bei diesen Systemen i. d. R. nicht um Online-Händler handeln, sondern um nicht-kommerzielle Literatur-Nachweissysteme, aber dennoch besteht die Möglichkeit, dass es zu manipulativen Kommentaren kommt. Es erscheint daher sinnvoll, sich über mögliche Gegenmaßnahmen Gedanken zu machen, auch wenn diese – wie Kneifel (2009, S. 134) anmerkt – aus Aufwandsgründen nur bis zu einem bestimmten Grad umsetzbar sind und die grundle-

---

gende Strategie in einem Vertrauensvorschuss in die Nutzergemeinschaft bestehen muss.

### 6.3 Fazit

Die Aufnahme kollaborativ erstellter Daten ist auf den ersten Blick eine einfache Erweiterung um zusätzliche Objekte. Die Personalisierung und – im Hinblick auf Kommentare – die möglichen Beziehungen dieser Objekte untereinander fügen jedoch eine Vielzahl neuer Gesichtspunkte und Anforderungen an den Anreicherungsdienst hinzu. Die daraus erwachsende Komplexität in der Datenhaltung / im Datenmodell, im Hinblick auf den Datenschutz sowie in der Gestaltung der Anzeige- und Bearbeitungsfunktionen legen den Schluss nahe, dass diese Funktionalität allenfalls in einem zweiten Schritt zu einem grundsätzlich funktionalen Dienst hinzugefügt werden können. Sollte sich bei näherer Analyse der Implementierungsaufwand als zu hoch erweisen, stellt eine kreative Nachnutzung vorhandener etablierter Dienste eine gute Alternative dar. Sie wäre allerdings stets mit dem Risiko verbunden, dass diese Dienste ihre Nutzungsbedingungen prohibitiv anpassen oder ihren Geschäftsbetrieb einstellen.

Die Herauslösung der sozialen Funktionen aus dem Kontext lokaler Anwendungen brächte in jedem Fall den Vorteil einer enorm ausgeweiteten potentiellen Benutzergemeinschaft. Wie Greenall (2011, Abschnitte „Commenting“ bzw. „Existing Approaches“) schreibt, ist es nur ein kleiner Teil der Benutzer, die aktiv die angebotenen Kommentarfunktionen verwendet. Diese Einschätzung wird durch die empirische Untersuchung von Kneifel (2009, S. 103 f.) auch hinsichtlich der Tagging-Funktion bestätigt. Ein großer Anteil der existierenden Werke ist jedoch in mehr als einem Katalog / mehr als einer Datenbank nachgewiesen. Durch die Nachnutzung von Kommentaren oder Tags über eine Austauschstelle steigt damit die Wahrscheinlichkeit, in einem Katalog auf ein solches Anreicherungs-element zu stoßen, stark an. Theoretisch könnte ein auf diese Weise ein globales Instrument zur Diskussion von Publikationen entstehen.

Es ist davon auszugehen, dass die großen Softwareanbieter OCLC und ExLibris entsprechende Funktionen in ihre derzeit in der Entwicklung befindlichen web- bzw. cloud-basierten Bibliothekssysteme Alma<sup>46</sup> bzw. Web-scale Management Services (WMS)<sup>47</sup> integrieren werden. Effektiv wird der Datenaustausch in diesen Fällen aber voraussichtlich auf die Kunden der entsprechenden Anbieter beschränkt bleiben und damit nicht sein volles Potential ausschöpfen können.

Ein unabhängiger Dienst würde es den Benutzern erlauben, ihre Tags gewissermaßen von einem System (Katalog oder Datenbank) zum anderen «mitzunehmen». Eine technische Herausforderung dürfte dabei das «Wiederfinden» von einmal getaggt oder kommentierten Werken sein, da es

---

<sup>46</sup> <http://www.exlibrisgroup.com/category/AlmaOverview>.

<sup>47</sup> <http://www.oclc.org/us/en/webscale/default.htm>.



---

zwar – entsprechend der Ausführungen im Abschnitt 6.1.4 – grundsätzlich umsetzbar ist, die für den jeweiligen Datenbestand vergebenen Tags für die Suche lokal zu importieren, für die Beschränkung auf die davon durch den jeweiligen Benutzer vergebenen Tags aber eine weitere Verweisebene eingezogen werden muss: Zuerst die Abfrage der vom jeweiligen Benutzer vergebenen Tags aus dem Anreicherungsdienst, dann die Prüfung, welche davon im aktuellen Kontext einen oder mehrere Treffer ergeben.

Die Erweiterung des skizzierten Anreicherungsdienstes um die in diesem Kapitel diskutierten Funktionalitäten wäre für Bibliotheken ein großer Schritt in Richtung «Web 2.0» im Sinne einer Austauschplattform von Informationen und Meinungen und unabhängig von der Realisierung von «CaESer» oder eines vergleichbaren Dienstes ein lohnender Gegenstand weiterer Untersuchungen.

---

## 7 LIZENZEN UND NUTZUNGSRECHTE

In diesem Bereich sind zwei Aspekte zu beleuchten. Zum einen ist zu klären, welche Daten durch den Anreicherungsdienst vorgehalten oder vermittelt werden können bzw. welche Rahmenbedingungen dazu einzuhalten sind. Zum anderen stellt sich die Frage nach den Nutzungsbedingungen für den Anreicherungsdienst und der von diesem bereitgestellten Datensammlung.

### 7.1 Rahmenbedingungen

Unterschieden werden muss zwischen den Metadaten («Ressource X hat eine Rezension Y») als Fakten und den Anreicherungsressourcen selbst. Einzelne Fakten sind nicht durch rechtliche Bestimmungen geschützt, Sammlungen von Fakten dagegen genießen durchaus gesetzlichen Schutz entweder – länderabhängig – im Sinne des Urheberrechts oder im Sinne rechtlicher Bestimmungen eigener Art («sui generis») für Datenbanken (vgl. Guide to Open Data Licensing).

Bestimmte Anreicherungsressourcen wie zum Beispiel Rezensionen oder (unter bestimmten Umständen) Abstracts unterliegen zudem unabhängig von der rechtlichen Stellung der Datensammlung, in der sie enthalten sind, dem Urheberrecht (vgl. B. Pohl, 2006, S. 235 ff. oder Hatcher, 2010, S. 5).<sup>48</sup>

Diese zweite Ebene rechtlicher Bestimmungen darf bei der Verwendung der Ressourcen keinesfalls übersehen werden.

Bei der Übernahme von Datensammlungen in die Datenbank des Anreicherungsdienstes und ebenso bei der Verwendung von Schnittstellen von weiteren Anbietern zur dynamischen Abfrage und anschließenden Weiterleitung von Anreicherungsressourcen sind die geltenden Benutzungsbedingungen zu beachten. Unabhängig von bzw. zusätzlich zu den vorstehend genannten rechtlichen Rahmenbedingungen umfasst dies bei der Nutzung von authentifizierungspflichtigen Schnittstellen auch noch die – implizit durch die Registrierung anzuerkennenden – vertraglichen Nutzungsbedingungen des Anbieters (vgl. Guide to Open Data Licensing).

Unter «Benutzungsbedingungen» sind dabei nicht zwangsläufig «Beschränkungen» zu verstehen, denn auch die Einräumung von Nutzungsrechten stellt eine Benutzungsbedingung dar. Heath und Bizer (2011, Abschnitt 4.3.3) fordern deshalb gerade jene Datenanbieter auf, die ihre Daten zur freien Benutzung und Übernahme freigeben wollen auf, diese Benutzungsbedingungen durch entsprechende Regularien formal zu dokumentieren. Dritte erhalten auf diesem Weg die Gewissheit, unter welchen Rahmenbedingungen sie die bereitgestellten Daten nutzen können und das sie dies in

---

<sup>48</sup> Für eine Beschreibung der rechtlichen Rahmenbedingungen von bibliographischen Daten vgl. auch Bérard (2011, S. 5 ff.).

---

rechtlich abgesicherter Form ohne Furcht vor Klagen oder nachträglich erhobenen Nutzungsgebühren tun können.

Bei allen weiteren Ausführungen ist zu beachten, dass sie allgemeinen Charakters sind und Sonderfälle nicht berücksichtigen können. Speziell bei der Verwendung von Daten, die nicht die verbreiteten Lizenzen oder Verzichtserklärungen von Creative Commons<sup>49</sup> oder Open Data Commons<sup>50</sup> nutzen, ist daher Vorsicht geboten:

There's no copyright law of the internet and what will often matter most for online open data will be the law of one country (and that country might not be your own). (Hatcher, 2010, S. 5)

Eine gründliche, insbesondere auch die rechtliche Situation in Deutschland erfassende Einführung in die Problematik inklusive verständlicher Erläuterungen bietet auch Kreutzer (2011), dessen Leitfaden zur weiteren Lektüre empfohlen wird.

## 7.2 Begriffsklärung: Lizenzen, Verzichtserklärungen, Normen

Im Linked Open Data Kontext werden zwei Mittel genutzt, um Rechte und Beschränkungen zu übertragen: Lizenzen und Verzichtserklärungen. Hinzu kommen (Gemeinschafts-) Normen zum Ausdruck von erwünschtem Verhalten. Die grundlegende Annahme hinter allen weiteren Ausführungen ist, dass die handelnde (juristische) Person die Rechte tatsächlich besitzt, die sie anderen einräumt oder untersagt. Die weiteren Aussagen in diesem Abschnitt stützen sich primär und soweit nicht anders angegeben auf Davis (2009) sowie auf Heath und Bizer (2011, Abschnitt 4.3.3).

Bei einer Lizenzierung überträgt eine Partei bestimmte Rechte an einem Werk oder einer Datenbank an eine andere Partei. Welche Rechte das sind und in welchem Umfang (exklusiv; für einen bestimmten Zeitraum; für einen bestimmten Zweck; ...) sie übertragen werden ist dabei nicht vorherbestimmt. Der Rechteinhaber kann die gleichen oder andere Rechte in gleichem oder abweichenden Umfang parallel auch an weitere Parteien übertragen.

Bei einer Verzichtserklärung (englisch: Waiver) dagegen gibt ein Rechteinhaber in der Regel alle Rechte an einem Werk oder einer Datenbank auf (bzw. verzichtet – da die Urheberrechte nicht aufgegeben werden können – auf die deren Durchsetzung). Zumindest bei den häufig genutzten Verzichtserklärungen CC0 oder Open Data Commons erfolgt die Aufgabe der Rechte auch unwiderruflich (vgl. Korn und Oppenheim, 2011, S. 5). Beliebige andere Parteien haben damit automatisch das Recht, das Werk oder die

---

<sup>49</sup> Siehe <http://creativecommons.org/> bzw. <http://de.creativecommons.org/>.

<sup>50</sup> Siehe <http://opendatacommons.org/>.

---

Datenbank nach eigenem Belieben zu nutzen und die Partei, die die Rechte aufgegeben hat, kann anschließend keine Kontrolle mehr darüber ausüben.

(Gemeinschafts-) Normen stellen in gewissem Rahmen ein Korrektiv für die absolute Aufgabe der Rechte dar. Mit ihnen kann der (ehemalige) Rechteinhaber versuchen, ein bestimmtes erwünschtes Verhalten bei der Verwendung von durch Verzichtserklärungen freigegebenen Werken/Datenbanken herbeizuführen. Durch die Aufgabe seiner Rechte verfügt er jedoch nicht mehr über juristische Mittel, die in den Normen formulierten Erwartungen durchzusetzen. Er kann lediglich hoffen, dass die Gemeinschaft – beispielsweise die bibliothekarische Fachöffentlichkeit – entsprechenden Druck auf solche Mitglieder ausübt, die gegen die Normen verstoßen.

In den vorstehend genannten Veröffentlichungen von Davis bzw. Heath/Bizer finden sich auch diverse Verweise auf nachnutzbare Lizenzen bzw. Verzichtserklärungen. Die Verwendung bestehender Lizenzen / Waiver ist gegenüber der Verwendung eigener Varianten zu bevorzugen, da damit für Interessenten die Notwendigkeit entfällt, unbekannte Texte individuell auf ihre Bedeutung hin untersuchen zu müssen, was je nach Komplexität der Bestimmungen die (teure) Konsultation von Experten erfordert.

In der Linked Data / Open Data Gemeinschaft wird grundsätzlich die Aufgabe aller Beschränkungen durch die Verwendung von Verzichtserklärungen angestrebt, da dies die Nachnutzbarkeit maximiert, während alle anderen Vorgehensweisen subtile Problematiken entfalten können (vgl. A. Pohl, 2010, S. 5 f.).

### 7.3 Konsequenzen für den Datenimport

Die Aufnahme von Datenbeständen in die zentrale Datenbank des Anreicherungsdienstes kann nur nach individueller Prüfung der mit dem jeweiligen Datenbestand verknüpften Bedingungen erfolgen. Zu beachten ist, dass Lizenzen untereinander inkompatibel sein können. Beispielsweise dürfen Daten, deren Lizenz eine kommerzielle Nutzung untersagt nicht in Kombination mit Daten weitergegeben werden, deren Lizenz dies erlaubt (vgl. Möller, 2005).

Die einzelnen Teile der aggregierten Datenbank unterliegen weiterhin den ursprünglichen Bedingungen, d. h. es ist selbstverständlich ausgeschlossen, die Datenbank des Anreicherungsdienstes insgesamt zum Beispiel unter Abgabe einer Verzichtserklärung freizugeben. Soweit sich aus einer rechtlich zulässigen Auswertung der aggregierten Teilbestände neue Erkenntnisse ableiten lassen (beispielsweise die Zuordnung von «verwandten» Identifikationsmerkmalen), spricht nichts gegen die Weitergabe dieser Daten.

Der Anreicherungsdienst ist primär auf die Nachnutzung von andernorts erstellten Ausgangsdaten angewiesen. Es ist daher angebracht und sinnvoll (im Sinne einer Bekräftigung des dahinter stehenden Ansatzes), solche Da-

---

ten, für die die notwendigen Rechte vorliegen, unter Anwendung einer Verzichtserklärung für jegliche Nachnutzung vollständig freizugeben.

Eine vollständige Auflistung möglicher Problemfälle ist aufgrund der Vielzahl an möglichen Lizenzmodellen und ihrer Wechselbeziehungen unmöglich. Daher werden nachstehend lediglich einige grobe Leitlinien aufgestellt.

Datenbanken mit Lizenzen, die eine kommerzielle Nutzung untersagen (speziell Creative Commons NC), sollten nicht integriert werden, da sie nicht mit Daten kombiniert werden können, die eine kommerzielle Nutzung erlauben, aber ein «Share Alike» (Weitergabe unter identischen Bedingungen) voraussetzen: beides schließt sich gegenseitig aus (vgl. Möller, 2005).

«Attribution»-Lizenzen, die eine Benennung des ursprünglichen Erzeugers des lizenzierten Objektes erzwingen, sind sehr verbreitet (vgl. Brown, 2004). Die Datenhaltung und die Schnittstellen des Anreicherungsdienstes müssen die zur Erfüllung der Auflage notwendigen Informationen unbedingt beinhalten, da sonst voraussichtlich ein großer Teil der interessanten Datenbestände nicht verwendet werden kann.

Unterschiedliche «Vertriebswege» (Datenexport / API) des Anreicherungsdienstes sind möglicherweise getrennt zu betrachten, falls eine Nutzungslizenz beispielsweise eine Weitergabe eines Datenbestandes in seiner Gesamtheit untersagt. Der Zugriff auf die API könnte von der Anerkennung bestimmter Nutzungsbedingungen abhängig gemacht werden. Inwieweit damit Beschränkungen durch restriktive Lizenzen vermieden werden könnten, hängt eventuell auch von den einzelnen Lizenzen ab und müsste separat untersucht werden. Die Frage kann daher an dieser Stelle nicht beantwortet werden.

Zu beachten ist noch, dass ein Datenlieferant sich zwar keine Rechte über gemeinfreie Werke anmaßen darf (also deren Verwendung in keiner Weise einschränken kann), sich aber durchaus für seinen Aufwand etwa bei der Digitalisierung, Zusammenstellung und Bereitstellung von Daten (also für seine Dienstleistung, nicht für die Daten an sich) entschädigen lassen darf.

## 7.4 Konsequenzen für dynamisch integrierte Quellen

Das theoretische Konzept des Anreicherungsdienstes sieht vor, auch Daten auszuliefern die sich nicht in der internen Datenbank des Dienstes befinden, sondern dynamisch auf Anfrage aus anderen Diensten abgerufen werden. Dazu ist es notwendig, auf Schnittstellen (API) des jeweiligen Anbieters zuzugreifen.

Sofern der Anbieter eine Registrierung im Vorfeld einer Nutzung der API voraussetzt (dies entspricht einem Vertragsabschluss) und nur authentifizierte Zugriffe auf die API zulässt, kann er über vertragliche Bestimmungen praktisch beliebige Nutzungsbedingungen durchsetzen (vgl. Guide to Open Data Licensing bzw. ausführlicher Kreutzer, 2011, S. 46 ff.). Die Spanne reicht von Beschränkungen der Zugriffe pro Zeiteinheit über erfor-

---

derliche Back-Links zum Anbieter bis hin zum Verbot eines Zwischenspeicherns oder der Vermittlung durch eine Zwischenstelle an sich. Die Freiheiten bei der Festlegung der Bestimmungen machen eine abstrakte Diskussion an dieser Stelle unmöglich. Standard-Lizenzen wie für Datensammlungen scheint es nicht zu geben. Klar ist daher, dass eine individuelle Prüfung der Bedingungen jedes einzelnen Anbieters unumgänglich ist.

## 7.5 Nutzungsbedingungen des Anreicherungsdienstes

Analog zu den ursprünglichen Datenlieferanten hat auch der Anreicherungsdienst (bzw. dessen Betreiber) das Recht, Bedingungen für die Nutzung seiner Schnittstellen festzulegen sowie sich für die Bereitstellung von Daten bezahlen zu lassen<sup>51</sup> (soweit ihm die Nutzungsbedingungen der Daten eine solche Bereitstellung erlauben). Während Nutzungsbedingungen für die Schnittstelle im Hinblick auf ein anwendungsfreundliches Verhalten unproblematisch sind, sollten im Hinblick auf die politische Zielsetzung des Dienstes keine Bedingungen formuliert werden, die dessen Verwendung einschränken könnten.

---

<sup>51</sup> Wie es beispielsweise die Deutsche Nationalbibliothek für ihre grundsätzlich gemeinfreien Anreicherungsdaten tut (vgl. Knull-Schlomann, 2009).

---

## 8 DAS PROBLEM DER IDENTITÄT

Typischerweise werden Anreicherungen im Katalog als Verweise an bestehende Datensätze angehängt oder anhand eines in diesem Datensatz vorkommenden Identifikators dynamisch zur Anzeige hinzugefügt.<sup>52</sup> Soweit eine Anreicherung nicht gezielt für einen bestimmten Datensatz erstellt wird<sup>53</sup>, erfolgt auch die Integration in den Datensatz i. d. R. über einen Abgleich zwischen Identifikatoren (meist in Form von Identifikationsnummern).

### 8.1 Fehlende Eindeutigkeit von Identifikatoren

Je nach verwendetem Identifikator ist dieser Abgleich mehr oder weniger präzise. Beispielsweise ist die Vorstellung falsch, dass eine ISBN eindeutig eine bestimmte Publikation identifiziert (Greenall, 2011, Abschnitt "Work-ID System"). In der Realität wurden ISBN durch Verlage absichtlich oder versehentlich für völlig voneinander unabhängige Publikationen wiederverwendet. Mit der Einführung der 13-stelligen ISBN kam noch die unzutreffende Annahme hinzu, dass ISBN13 mit dem Präfix «978» losgelöst von den bisherigen ISBN10 zu betrachten seien, was zur erneuten Vergabe bereits verwendeter ISBN10 in ISBN13-Schreibweise führte.

Abhängig von Qualitätsansprüchen und Aufwand werden daher mancherorts Plausibilitätsprüfungen in Anreicherungsprozesse eingebaut. Diese erfordern allerdings, dass die Anreicherungsressource selbst wiederum mindestens einen rudimentären Satz an Metadaten enthält, damit ein erweiterter Abgleich stattfinden kann.

Andere Identifikatoren sind – nicht zuletzt wohl aufgrund einer späteren Entwicklung in einem anderen, den damit verknüpften Problemen gegenüber aufgeschlosseneren Umfeld sowie einer strikteren Definition – glücklicherweise präziser.<sup>54</sup>

### 8.2 Tolerante Verwendung von Identifikatoren

Einerseits ist die präzise Zuordnung von Anreicherungsressourcen zu bibliographischen Ressourcen also erwünscht, weil es darum geht, nicht zugehörige Angaben zu vermeiden, die Katalogbenutzer irreführen würden. Andererseits ist eine gewisse Toleranz in diesem Kontext jedoch wünschenswert.

---

<sup>52</sup> Für Details vgl. Kapitel 2.4.

<sup>53</sup> Z. B. indem während des Katalogisierungsprozesses das Inhaltsverzeichnis gescannt wird.

<sup>54</sup> Im Unterschied zu vielen Fällen bei ISBN ist dem Autor in der beruflichen Praxis bisher kein Fall einer doppelt vergebenen DOI begegnet – was deren Existenz keineswegs ausschließt, aber zumindest einen Hinweis auf eine präzisere Vergabepaxis darstellt.

---

Was zunächst paradox klingt, erklärt sich einfach durch die Tatsache, dass es um eine (theoretisch) klar definierte Toleranz handelt. Es geht dabei keinesfalls darum, die Präzision der Identifikation von bibliographischen Ressourcen durch Identifikatoren zu verringern. Sondern es geht darum, bei der Zuordnung von Anreicherungen zu bibliographischen Ressourcen ein anderes Verständnis von Äquivalenz anzuwenden.

Schlagwörter, Abstracts oder Rezensionen zu einem Titel verlieren beispielsweise nicht oder nur in geringem Maß an Relevanz, wenn sie im Kontext einer anderen Auflage oder einer Übersetzung des gleichen Werkes angezeigt werden. Erkennt und in gewissem Rahmen umgesetzt hat dies etwa Amazon: dort werden – wie in Abbildung 2 zu sehen<sup>55</sup> – Kommentare für andere Ausgaben (Taschenbuch / gebundene Ausgabe) ebenfalls angezeigt.

### 8.3 Bestimmung von Toleranzgrenzen

Zu hinterfragen ist allerdings, ob diese Möglichkeit der Übernahme von Anreicherungen von ihrem ursprünglichen Bezugsobjekt auf andere Bezugsobjekte grundsätzlich gilt, oder ob sie einerseits vom Typ der Anreicherungsressource und andererseits vom Typ der Beziehung zwischen den bibliographischen Ressourcen abhängt.

Ein sachlich orientiertes Schlagwort oder Tag stellt i. d. R. eine relativ grobe Beschreibung eines Titels dar.

---

<sup>55</sup> Vgl. [http://www.amazon.de/product-reviews/3551551677/ref=cm\\_cr\\_pr\\_top\\_recent?ie=UTF8&showViewpoints=0&sortBy=bySubmissionDateDescending](http://www.amazon.de/product-reviews/3551551677/ref=cm_cr_pr_top_recent?ie=UTF8&showViewpoints=0&sortBy=bySubmissionDateDescending); Ausgabe in Abbildung 2 wurde gekürzt.



The screenshot shows the Amazon.de interface for the product 'Harry Potter und der Stein der Weisen (Band 1)'. The top navigation bar includes the Amazon logo, user account links, and search options. The main content area displays the product title, a star rating of 4.5 (based on 802 reviews), and a price of EUR 15,90. A sidebar on the right shows 'Dieses Produkt' with a book cover and 'Kunden, die diesen Artikel angesehen haben, haben auch angesehen' with other book covers. The bottom section shows customer reviews with star ratings and text.

Abbildung 2 Amazon: Kunden-Rezensionen zu anderer Ausgabe

Selbst eine grundlegende Überarbeitung eines Titels im Rahmen einer Neuauflage wird kaum eine so gravierende Änderung der behandelten Thematik mit sich bringen, dass ein solches Schlagwort den Inhalt nicht mehr zutreffend benennt. Eine Rezension dagegen kann sich durchaus zu erheblichen Anteilen auf die Kritik an Schwachpunkten konzentrieren, die durch eine Bearbeitung und Neuauflage ausgeräumt werden und damit ihre Relevanz in großen Teilen einbüßen. Dies gilt aber nicht generell für Neuauflagen, sondern eben nur für in bestimmten Punkten überarbeitete Neuauflagen. Derartige Änderungen sind jedoch nicht anhand von Metadaten erkennbar, sondern erfordern eine intellektuelle Analyse der Texte: Selbst wenn die Auflagenbezeichnung den häufig gebrauchte Ausdruck «grundlegend überarbeitet» enthält, lässt sich daraus keinerlei Bezug zu konkreten Kritikpunkten ableiten.

Umschlagabbildungen fallen intuitiv eher in die Kategorie der nicht sinnvoll übertragbaren Anreicherungsdaten, weil der Wiedererkennungswert bei stark abweichende Gestaltung zwischen unterschiedlichen Ausgaben oder gar zwischen der DVD-Hülle einer Verfilmung und einer gedruckten Ausgabe nicht mehr gegeben ist. O'Connor und O'Connor (1998) beschreiben andererseits, dass – neben Klappentexten – auch die Abbildungen auf (Schutz-) Umschlägen spezifische Informationen zu einem Buch oder anderem Medium transportieren und eine wichtige Rolle bei der Vorselektion von Titeln spielen. Unter dem Vorbehalt, dass sich die durch die Gestaltung vermittelten Informationen nicht «widersprechen», kann also ein zusätzli-

ches Umschlagbild als ein Zugewinn an Informationen für den Katalogbenutzer verstanden werden. Ein anderes Argument für die Aufnahme nicht direkt zuzuordnender Abbildungen könnte sein, dass ein Katalogbenutzer positive Erfahrungen mit Titeln aus einer Reihe gemacht hat, die aktuell angezeigte Auflage des Titel zwar in eine abweichend gestaltete Übersetzung der Reihe gehört aber durch die Anzeige der «verwandten» Umschlagbilder trotzdem intuitiv mit den bekannten Titeln in Zusammenhang gebracht werden kann. LibraryThing jedenfalls zeigt auf seinen «Werk»-Seiten eine Vielzahl von Umschlagbildern an, wie Abbildung 3 zeigt.<sup>56</sup>

**LibraryThing** Sign in / Join English Translate! Help

Home Zeitgeist Talk Groups Local

Search site

**Adventures of Huckleberry Finn**  
by Mark Twain

Series: Tom Sawyer (2)

Members	Reviews	Popularity	Average rating	Conversations
16,723	243	66	★★★★ (3.96)	514

Member-uploaded covers (438)

**Quick Links**

- Amazon.com (direct)
- Abebooks.com
- Google Books — Partial view
- Project Gutenberg (2 editions)
- WorldCat

**Get this book**

- Local Book Search
- All sources

**Swap**  
779 avail. 320 wanted

**Ebooks**  
9 free 40 pay

**Audio**  
4 free 30 pay

**Popular covers**

(see all 673 covers)

**Penguin Australia**

Eight editions of this book were published by Penguin Australia.

Editions: 0140424385, 0141439645, 0141023619, 0141321091, 0451530942, 0141045183, 0143105949, 0141334843

Abbildung 3 Umschlagbilder bei LibraryThing

Wie relevant welches Argument in der Praxis ist, müsste im einzelnen empirisch untersucht werden – die Ausführungen hier sollten jedoch zeigen, dass die Entscheidung über die Anzeige oder Nicht-Anzeige bestimmter Anreicherungsdaten keineswegs trivial sind. In der Praxis wird man sich für einen formal / algorithmisch aus den Metadaten ableitbaren Weg entscheiden und dessen Auswirkungen – in der Form von Irritationen oder Zusatzin-

<sup>56</sup> <http://www.librarything.com/work/3093889/covers/>.

---

formationen für die Katalogbenutzer – messen müssen.<sup>57</sup> Die Ergebnisse der Tests können anschließend in Änderungen des Algorithmus einfließen. Grundsätzlich ist zu konstatieren, dass es in diesem Kontext nie um «100%»-Lösungen geht, sondern stets um eine Optimierung der Trefferquote.

## 8.4 Functional Requirements for Bibliographic Records (FRBR)

Um die Zuordnung von Anreicherungsdaten zu anderen als den ursprünglich als Bezugsobjekt festgelegten bibliographischen Ressourcen überhaupt formal festlegen zu können, ist ein theoretisches Modell der Beziehungen zwischen bibliographischen Ressourcen notwendig. Ein solches findet sich als Teil der sogenannten «Functional Requirements for Bibliographic Records» (FRBR)<sup>58</sup>.

### 8.4.1 Entitätenmodell

Während sich eine praktische Relevanz der FRBR als Gesamtheit auch über 10 Jahre nach ihrer Publikation nur bedingt entfaltet hat (Wiesenmüller, 2008, S. 348 f. oder Dickey, 2008, S. 25 ff.), erfreut sich zumindest das Entitätenmodell zur Gruppierung der verschiedenen «Realisierungen» eines Werkes bzw. zum Ausdruck von Beziehungen zwischen bibliographischen Ressourcen (Dickey, 2008, S. 23) erheblicher Popularität.<sup>59</sup> Dies liegt vermutlich nicht zuletzt daran, dass die grundsätzlichen Ideen dazu bereits erheblich länger existieren, wenngleich sie zuvor nicht in vergleichbarem Rahmen formalisiert wurden.<sup>60</sup>

Das Modell definiert in der sogenannten Gruppe 1 vier Entitäten für «die Produkte von intellektuellen bzw. künstlerischen Anstrengungen, die in bibliografischen Sätzen genannt oder beschrieben sind» (IFLA, 2009, S. 11): Werk (work), Expression (expression), Manifestation (manifestation) und Exemplar (item). Abbildung 4 stellt diese Entitäten und ihre grundlegenden Beziehungen im Überblick vor (Wiesenmüller, 2008, S. 350).

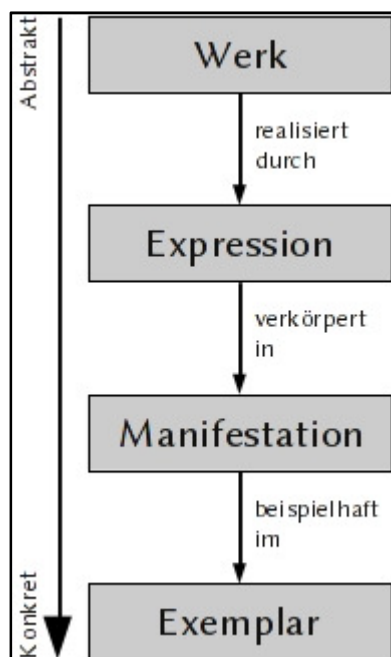
---

<sup>57</sup> Beispielsweise könnten neben den Anreicherungsressourcen jeweils zwei Bedienelemente «Diese Information war hilfreich» bzw. «Diese Information war irritierend» angezeigt werden, über die die Benutzer mit minimalem Aufwand Feedback geben könnten. Beispielsweise mit Hilfe eines A/B-Tests (vgl. [http://en.wikipedia.org/wiki/A%2FB\\_testing](http://en.wikipedia.org/wiki/A%2FB_testing)), der eine tolerantere und eine weniger tolerante Methode anhand des eingehenden Feedback vergleichen würde, könnte der Zusammenführungsalgorithmus schrittweise optimiert werden.

<sup>58</sup> IFLA, 2009.

<sup>59</sup> Ein Hinweis auf die mangelnde Verankerung des FRBR-Konzeptes im Bewusstsein der Fachöffentlichkeit mag auch die Tatsache sein, dass Hauke, Hötzel und Rumler (2010) kein einziges Mal darauf Bezug nehmen, obwohl sie sich in ihrem Artikel exakt mit jenen Problemen beschäftigen, die durch FRBR gelöst werden sollen.

<sup>60</sup> Vgl. die historische Betrachtung zur Entwicklung durch Denton (2007).



**Abbildung 4 FRBR-Entitäten für bibliografische Ressourcen**

Das Werk ist die abstrakteste der vier Entitäten; es ist definiert als eine «individuelle intellektuelle bzw. künstlerische Schöpfung» (IFLA, 2009, S. 16). Damit ist nicht die Erstveröffentlichung gemeint, sondern eher etwas wie die vollständige Idee oder Vorstellung eines Autors über eine Geschichte oder das Bezugsobjekt einer Person, die beispielsweise allgemein über den «Herrn der Ringe» spricht, ohne die englische Buchfassung, die erste oder zweite deutsche Übersetzung oder den Film zu adressieren (vgl. Wiesenmüller, 2008, S. 350).

Die nächst-konkretere Entität ist die Expression, die definiert ist als «die intellektuelle bzw. künstlerische Realisierung eines Werkes in Form von Buchstaben, Zahlen, Noten [...] usw. oder einer Kombination dieser Formen» (IFLA, 2009, S. 18). Beispiele für verschiedene Expressionen eines Werkes sind ein Roman und dessen Übersetzung in eine andere Sprache ebenso wie eine Neu-Übersetzung in die gleiche Sprache. Überarbeitungen einer Expression, die über ein geringes Maß (im Sinne etwa von Rechtschreibkorrekturen) hinausgehen, erzeugen eine neue Expression (vgl. Wiesenmüller, 2008, S. 350) – es sei denn, sie beinhalten ein signifikantes Maß «an unabhängiger künstlerisch-intellektueller Schöpfung: In einem solchen Fall entsteht keine neue Expression, sondern ein neues, verwandtes Werk» (Wiesenmüller, 2008, S. 350).

---

Die dritte, wiederum konkretere Entität ist die Manifestation, die «physische Verkörperung einer Expression eines Werkes» (IFLA, 2009, S. 21).<sup>61</sup> Diese Entität entspricht in etwa einer Ausgabe (die bislang das primäre Katalogisierungsobjekt darstellte), ist aber nicht damit gleichzusetzen, da eine unveränderte Neuauflage keine neue Manifestation darstellt (Wiesenmüller, 2008, S. 351). Änderungen im Erscheinungsbild (beispielsweise im Seitenlayout oder in der Schriftart), dem Trägermedium oder des Verlegers führen damit zu einer neuen Manifestation.

Abschließend folgt als letzte Entität das Exemplar als einzelnes Stück einer Manifestation (das durchaus aus mehreren Bestandteilen – beispielsweise Bänden – bestehen kann) (IFLA, 2009, S. 23). Grundsätzlich sind alle Stücke identisch, aber Änderung nach der eigentlichen Herstellung (Buchbindearbeiten, Randnotizen o.ä.) können zu individuellen Eigenschaften führen. Da es um den Aufbau eines zentralen Anreicherungsdienstes für viele Anwender geht, Exemplare jedoch per Definition Einzelstücke sind und spezifische Informationen über ein Exemplar somit nicht auf andere übertragen werden können, ist eine Berücksichtigung dieser Ebene im Kontext der vorliegenden Arbeit überflüssig.

#### 8.4.2 Beziehungen zwischen Entitäten

Wie in den Abschnitten 8.2 und 8.3 dargelegt, wäre es für eine Kataloganreicherung sehr hilfreich, wenn bibliographische Ressourcen gruppiert werden könnten, um die den einzelnen Mitgliedern der Gruppe zugeordneten Anreicherungen im Sinne einer transitiven Übertragung an die übrigen Mitglieder durchzuschleifen. Die Zusammenfassung von Ressourcen unter dem Dach eines FRBR-Werkes wäre dabei offensichtlich bereits sehr hilfreich. Andererseits ist auch die Übertragung bestimmter Anreicherungsressourcen von einem Werk auf ein anderes vorstellbar (beispielsweise von Schlagwörtern von einer Fassung eines Textes für Erwachsene zu einer Fassung für Kinder), während eine Rezension für eine bestimmte Manifestation viel, für eine überarbeitete andere Manifestation dagegen kaum Sinn machen kann.

Das FRBR-Konzept enthält einen weiteren Bestandteil, der zu einer Verbesserung der Situation führen kann: Beziehungen zwischen Entitäten bzw. zwischen Instanzen dieser Entitäten (vgl. IFLA, 2009, S. 55 ff.). Primär von Interesse sind im vorliegenden Kontext die Beziehungen zwischen Entitäten der Gruppe 1 (vgl. vorhergehenden Abschnitt), die es eventuell ermögli-

---

<sup>61</sup> Der explizite Bezug auf eine physische Dimension erscheint im Hinblick auf die Existenz von rein digitalen E-Books merk- und fragwürdig. Leider geht der FRBR-Bericht ebenso wenig auf diesen Punkt ein, wie der bereits mehrfach zitierte Artikel von Wiesenmüller, obwohl eines der im Bericht definierten Merkmale für eine Manifestation eine Zugangsadresse für eine elektronische Ressource im Fernzugriff ist (vgl. IFLA, 2009, S. 48).

---

chen, gezielt bestimmte Anreicherungsressourcen an bestimmte bibliographische Ressourcen zu übertragen.<sup>62</sup>

Ignoriert man die Beschreibung von Beziehungen, in denen Exemplare beteiligt sind<sup>63</sup>, sowie allgemeine Erläuterungen verbleiben immerhin 16 Seiten zur Erläuterung von Beziehungen auf höheren Ebenen (vgl. Kapitel 5.3.1 bis 5.3.4 in IFLA, 2009, S. 64–80):

- Werk-zu-Werk-Beziehungen (differenziert in Beziehungen zwischen voneinander abhängigen und unabhängigen Werken)
- Beziehungen zwischen dem Ganzen und den Teilen auf der Werk-Ebene
- Expression-zu-Expression-Beziehungen (differenziert in Beziehungen zwischen Expressionen des gleichen und unterschiedlichen Werken und letzteres wiederum differenziert nach voneinander abhängigen und unabhängigen Werken)
- Beziehungen zwischen dem Ganzen und den Teilen auf der Expression-Ebene
- Expression-zu-Werk-Beziehungen (zwischen Expressionen eines Werkes und anderen Werken)
- Manifestation-zu-Manifestation-Beziehungen
- Beziehungen zwischen dem Ganzen und den Teilen auf der Manifestation-Ebene

Um Entscheidungen über den Transfer von Anreicherungen von einer Seite eine Beziehung zur anderen Seite – und letztlich von dort über weitere Beziehungen zu wiederum weiteren Entitäten – treffen zu können, müssten alle von der FRBR-Studie vorgesehenen Beziehungen analysiert werden. Bereits in der Studie werden diese Beziehungen weiter qualifiziert. Beispielsweise werden Beziehungen zwischen abhängigen Werken und unabhängigen Werken unterschieden, wobei etwa eine Fortsetzung eines Werkes beides sein kann (IFLA, 2009, S. 64), also eine intellektuelle Entscheidung erfordert. Zusätzlich müsste eine Matrix aufgespannt werden, die auch noch den Typ der Anreicherung einbezieht: Je konkreter der Bezug einer Anreicherungsressource auf eine bestimmte Entität und ihr Detailgrad sind, desto unwahrscheinlicher wird eine Übertragbarkeit. Ein Schlagwort wird voraus-

---

<sup>62</sup> Anreicherungen z. B. in Form von Verweisen auf die persönlichen Websites von Autoren berühren Beziehungen zwischen Entitäten verschiedener Gruppen, sind aber für die Fragestellung nach der Übertragbarkeit von Anreicherungen nicht von Belang. Bestimmte Anreicherungen – etwa die Angabe einer Rezension zu einer bibliographischen Ressource – entsprechen einer Beziehung zwischen Entitäten der Gruppen 1 und 3 (vgl. Wiesenmüller, 2008, S. 353 und S. 354), obwohl es sich vordergründig um Beziehungen innerhalb der Gruppe 1 handelt. Auch diese Art von Beziehungen ist im vorliegenden Kontext nicht von Bedeutung.

<sup>63</sup> Wie bereits zuvor beschrieben, macht es in einen institutions-übergreifend genutzten Dienst keinen Sinn, spezifische Daten zu Objekten zu sammeln, die per Definition einzigartig sind.

---

sichtlich auch jeder anderen Expression des gleichen Werkes zugeordnet werden können, während eine Inhaltsangabe zu einer gekürzten Fassung für eine erweiterte Fassung nicht mehr angemessen sein muss.

#### 8.4.3 Praktische Anwendbarkeit

Insgesamt bietet das FRBR-Modell das Potential, das Benutzungserlebnis eines Katalogs für dessen Benutzer erheblich zu verbessern, wie Denton (2006) beispielhaft zeigt. Wiesenmüller (2008, S. 358) wiederum weist explizit daraufhin, dass eine «FRBRisierung» von Katalogen durch die Zusammenführung von Daten zwischen Manifestationen eine erhebliche Verbesserung der Sacherschließung zur Folge haben würde. Auch OCLC betrachtet die Gruppierung von bibliographischen Ressourcen anhand des FRBR-Konzeptes zur Übernahme von Anreicherungen für alle Objekte innerhalb der Gruppe als erfolgversprechendes Mittel zur Optimierung von Katalogen im Hinblick auf die Bedürfnisse der Benutzer (vgl. OCLC Online Computer Library Center, Inc. 2009, S. 53).

Schwierigkeiten bei der Umsetzung des Modells in die Praxis der Katalogisierung können einerseits aufgrund der teilweise unscharfen und auslegungsbedürftigen Abgrenzung der Entitäten entstehen und andererseits aus unzureichenden Informationen des Katalogisierers über die Entstehungsgeschichte des ihm jeweils vorliegenden Objekts (vgl. Wiesenmüller, 2008, S. 350 f.) resultieren (insbesondere im Hinblick auf die Expression-Entität).

Wenn «[...] die Beziehungen im Kontext der für das Modell definierten Entitäten untersucht [werden]» (IFLA, 2009, S. 55), dann ist damit die implizite Aussage verknüpft, dass alle zu verknüpfenden bibliographischen Ressourcen analysiert und in entsprechende Entitäten aufgeteilt wurden, die wiederum für sich genommen eindeutig adressierbar sein müssen. Informelle Beziehungen über ad hoc aufteilbare Datenfelder innerhalb eines Datensatzes (vgl. Wiesenmüller, 2008, S. 353) genügen nicht. Die Adressierung muss zudem automatisiert verarbeitet werden können: Freitext-Bemerkungen (z. B. in Fußnoten) reichen nicht, selbst wenn das Werk darin benannt wird (vgl. IFLA, 2009, S. 55 f.).

Eine Frage die sich in dieser Hinsicht (letztlich auch unabhängig vom eigentlichen FRBR-Konzept) stellt ist, welche Bedeutung die Gruppierung von Ressourcen in der Praxis hat. Bates (2003, S. 39 f.) geht davon aus, dass die überwiegende Mehrheit der bibliographischen Ressourcen für sich allein steht. Wiesenmüller (2008, S. 357) führt ihrerseits Untersuchungen an, die diese Einschätzung belegen. Insofern erscheint der Aufwand für eine Zuordnung zu Entitäten und anschließende Gruppierung möglicherweise nicht gerechtfertigt. Andererseits kann man, so Bates (2003, S. 40) weiter, davon ausgehen, dass die Ressourcen, die zu großen Gruppierungen zusammengeführt werden können, eben auch solche Werke repräsentieren, die eine besondere Bedeutung haben und auf entsprechendes Interesse stoßen.

Alle Untersuchungen und Einschätzungen zur Anwendbarkeit, Umsetzbarkeit und tatsächlichen Effizienz basieren jedoch auf einem relativ kleinen



---

Schatz an Erfahrungen. Wiesenmüller (2008, S. 358) ebenso wie aktuell Coyle (2011) weisen daraufhin, dass es nur wenige «FRBRisierte» Datenbestände gibt – und von diesen existierten oder existieren wiederum einige nur in experimentellen Umgebungen. Zusätzlich fanden die Untersuchungen des FRBR-Modells offenbar jeweils auf Basis von zusammengehörigen – und dementsprechend zumindest grundlegend harmonisierten – Datenbeständen statt. Eine solche Datenbasis existiert für den projektierten Anreicherungsdienst jedoch nicht.

Die Bürde, neben der Implementierung der Datensammlung, Datenhaltung und Datenbereitstellung auch noch eine «FRBRisierung» vorzunehmen, kann einem Anreicherungsdienst nicht auferlegt werden, wenn dieser eine Aussicht auf Realisierbarkeit behalten soll.

Dies ist keinesfalls gleichbedeutend mit der Aussage, dass eine solche Aufgabe unlösbar wäre. Tatsächlich verwendet LibraryThing erfolgreich die Konzepte «Werk» und «Edition» zur Gruppierung von Einträgen in seiner Datenbank und kommt damit diesem Aspekt des FRBR-Modells sehr nahe (Czerwinski und Voß, 2010, S. 337). Die Identifikation und Gruppierung der Entitäten beruht allerdings zu einem erheblichen Anteil auf manueller Pflege durch die Benutzer von LibraryThing (Czerwinski und Voß, 2010, S. 337). Etwas vergleichbares ließe sich nur realisieren, wenn es gelänge, eine unmittelbar für Benutzer relevante Dienstleistung darum herum zu etablieren, was jenseits der in dieser Arbeit angenommenen Rahmenbedingungen läge. Die von LibraryThing selbst gesammelten Daten können nur eingeschränkt genutzt werden, da sie das Geschäftskapital der Firma darstellen und nicht frei zur Verfügung stehen.

## 8.5 Alternativen zur Gruppierung von bibliographischen Ressourcen

Während eine Implementierung des Entitäten- und Beziehungsmodells der «Functional Requirements for Bibliographic Records» aufgrund des dazu notwendigen Aufwands leider nicht in Frage kommt, bleiben die potentiellen Vorteile, also die stärkere Ausnutzung der vorhandenen Anreicherungsdaten durch eine Übertragung auf «verwandte» bibliographische Ressourcen unbestritten. Somit stellt sich die Frage, ob es eine konzeptionell ähnliche Alternative gibt, die aber mit einer niedrigeren Einstiegshürde verbunden wäre.

Einen Ausgangspunkt für eine solche Alternative bieten einige frei oder eingeschränkt verfügbare Datenbestände. Zum Ersten stellt LibraryThing einen Export der erfassten ISBN bereit, bei dem die ISBN nach den bei LibraryThing definierten Werken gruppiert sind (Spalding, 2007).<sup>64</sup>

---

<sup>64</sup> Der letzte verfügbare Export datiert allerdings vom 27.01.2009; vgl. <http://www.librarything.com/feeds/>.



Zum Zweiten sind die Daten des Open Library Projektes<sup>65</sup> frei verfügbar und enthalten sowohl Gruppierungen nach «Werken» als auch – innerhalb einzelner «Editionen» – eine Reihe von Identifikatoren, wie Abbildung 5<sup>66</sup> zeigt.

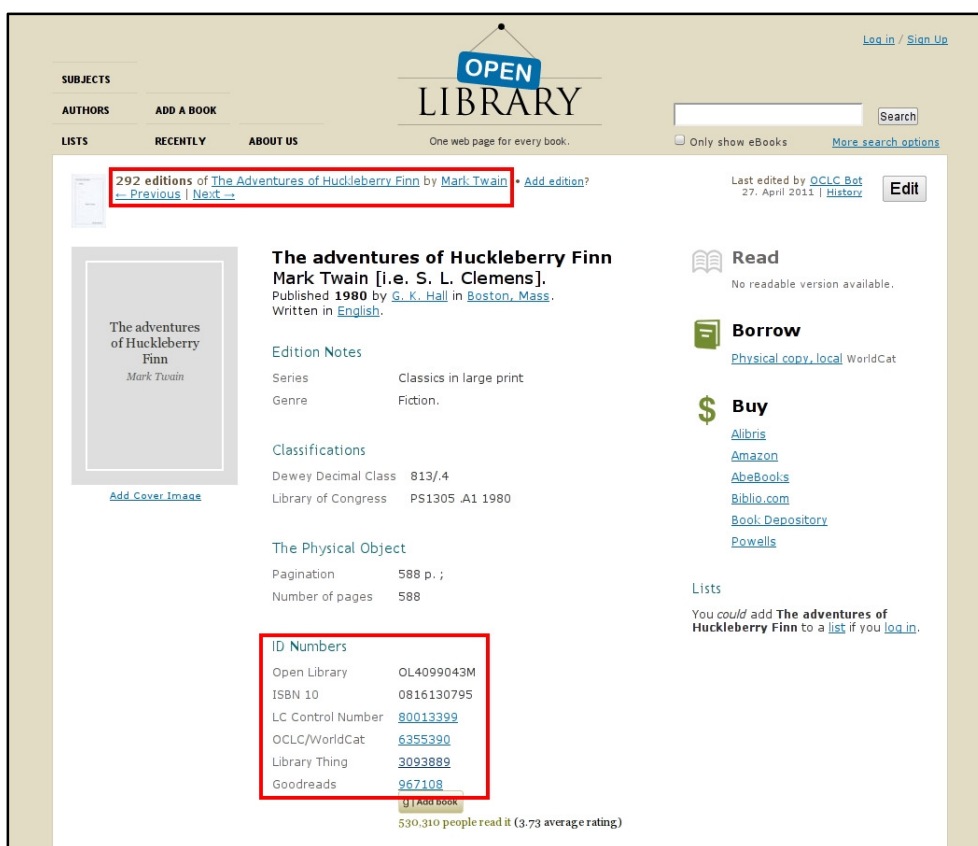


Abbildung 5 Edition mit Identifikatoren bei Open Library

Schließlich bietet OCLC mit seinen Diensten xISBN<sup>67</sup> und xOCLCNUM<sup>68</sup> die Möglichkeit, von ISBN zu OCLC-Nummern und von dort wiederum zu OCLC WorkID zu gelangen, wobei zu der Werk-ID wiederum zugehörige weitere OCLC-Nummern geliefert werden können. Die Abfrage dieses API ist zudem auch mit LCCN möglich. Allerdings ist die freie Nutzung auf 1000 Anfragen pro Tag limitiert<sup>69</sup>, so dass eine Nutzung im Kontext eines zentralisierten Anreicherungsdienstes für viele Nutzer-Institutionen den Erwerb einer kommerziellen Lizenz voraussetzen würde. Greenall (2011, Abschnitt "Work-ID System") hat diese Schnittstellen er-

<sup>65</sup> <http://openlibrary.org/>.

<sup>66</sup> [http://openlibrary.org/books/OL4099043M/The adventures of Huckleberry Finn](http://openlibrary.org/books/OL4099043M/The_adventures_of_Huckleberry_Finn).

<sup>67</sup> <http://xisbn.worldcat.org/xisbnadmin/doc/api.htm>.

<sup>68</sup> <http://xisbn.worldcat.org/xisbnadmin/xoclnum/index.htm>.

<sup>69</sup> Vgl. <http://xisbn.worldcat.org/xisbnadmin/index.htm>.

---

folgreich in seinen Prototypen eines kollaborativen, verteilten Kommentierungssystems eingebunden.

Darüber hinaus können frei verfügbare Datenbestände, etwa in Form von unter freie Lizenzen gestellten Metadaten von Bibliothekskatalogen, einerseits auf vorhandene Anreicherungsressourcen untersucht werden und zugleich die enthaltenen Identifikatoren extrahiert werden. Identifikatoren aus einem Datensatz wären grundsätzlich als gleichbedeutend zu betrachten, mit Ausnahme solcher, die als Verweise auf andere Datensätze dienen. Ob letztere weitere Identifikatoren zu einer vorhandenen Gruppe hinzufügen oder separat zu speichern sind, kann nur nach Analyse des jeweiligen Datenbestandes entschieden werden. Tritt ein bereits bekannter Identifikator (z. B. eine ISBN aus den LibraryThing Daten) erneut auf, können die weiteren Identifikatoren aus dem gleichen Datensatz zu einer schon vorhandenen Gruppierung hinzugefügt werden.

Zudem besteht die Hoffnung, Daten aus Projekten wie <sameAs><sup>70</sup> oder – zukünftig – insbesondere Culturegraph<sup>71</sup> übernehmen zu können.

Die Erfahrung muss zeigen, ob die Mischung der unterschiedlich tief greifenden Zusammenführungen bzw. das Fehlen von qualifizierten Beziehungen beispielsweise in den LibraryThing ISBN-Gruppen ein Problem darstellt und zu verwirrenden Anzeigen führt. Falls ja und falls die verfügbaren Daten weiterhin nicht für den Aufbau von FRBR-Entitäten und deren Beziehungen ausreichen, könnte eine Konsequenz in der grundsätzlichen Reduktion auf das Minimalmodell bestehen (oder alternativ darin, mehr Ressourcen in den Ausbau der Verknüpfung zu investieren – sofern die Ressourcen verfügbar sind).

---

<sup>70</sup> <http://sameas.org/about.php>.

<sup>71</sup> [http://www.culturegraph.org/resources/static/about\\_de.htm](http://www.culturegraph.org/resources/static/about_de.htm).

---

## 9 DATENHALTUNG UND DATENMODELL

Die Ausführungen in den Kapiteln 5 und 8 zeigen, dass der zu entwickelnde Dienst komplexe Datenstrukturen erfordert. Die verschiedenen Typen von Anreicherungsressourcen, die bereitgestellt werden könnten, lassen sich nicht a priori abschließend aufzählen, d. h. es muss möglich sein, jederzeit zusätzliche Angaben zuvor unbekannter Art für existierende und neue Identifikatoren ergänzen zu können. Eine einfache Tabellen-Struktur wie in Tabelle 1 ließe sich zwar leicht durch neue Spalten erweitern, ist aber aufgrund der spärlich gefüllten Zellen ineffizient und stößt an seine Grenzen, sobald mehr als eine Anreicherungsressource des gleichen Typs (z. B. URL zum Umschlagbild) sauber abgelegt werden soll.

ID	Cover	Abstract	Schlagwort	...
ISBN 0-12345-678-0			Programmierung	
...				

**Tabelle 1** (Unzureichende) Einfache Tabellen-Struktur

Entweder muss also ein komplexeres Schema für eine SQL-Datenbank entwickelt, oder die Wahl muss auf eine andere Datenbanktechnik fallen. Während aktuell eine Vielzahl an Datenbanktypen miteinander wetteifert<sup>72</sup>, erfreut sich in der jüngeren Vergangenheit nicht nur im bibliothekarischen Kontext<sup>73</sup> das Thema RDF wachsender Aufmerksamkeit und in dessen Folge das Konzept spezieller Datenbanken für RDF-Daten, die allgemein als «Triple Stores» bezeichnet werden. Graves, Constabaris und Brickley (2007, S. 196) betonen die besondere Eignung von RDF zur Zusammenführung von verteilten Datenbeständen und die Möglichkeit, mit einer einzigen Abfrage-Methodik auf dem gesamten Bestand operieren zu können, so dass ein näherer Blick sinnvoll erscheint.

### 9.1 Exkurs: Semantic Web, Linked (Open) Data und RDF

Die Vision des «Semantic Web» wurde bereits vor 10 Jahren von Tim Berners-Lee in einem Artikel dargelegt (Berners-Lee, Handler und Lassila, 2001). Reduziert auf das Kern-Konzept geht es darum, als evolutionäre Weiterentwicklung des WWW nicht länger nur von Dokumenten auf andere Dokumente zu verweisen, sondern auf der Ebene einzelner Informationsobjekte typisierte – also mit einer bestimmten Bedeutung versehene – Bezie-

---

<sup>72</sup> In der Diskussion findet sich häufig der Begriff «NoSQL» («Not only SQL»), der als Einstiegshilfe in das Thema dienen mag.

<sup>73</sup> Gradmann (2005) empfahl bereits vor ca. 6 Jahren die Adaption von RDF (und FRBR) als technologische Basis bibliographischer Datenhaltung.

---

hungen zu anderen Informationsobjekten herzustellen. Diese Verknüpfungen und ihre Bedeutungen sollen zudem maschinell verfolgbar und interpretierbar sein, um die Isolation einzelner Informationen aufzuheben, ausgehend von einem Startpunkt zu vielen weiteren direkt oder indirekt verbundenen Informationen zu gelangen und aufgrund eines «Verständnisses» der Beziehungen im Sinne von ableitbaren Schlussfolgerungen zielgerichtete Aktionen zu ermöglichen.

The Semantic Web is not a separate Web but an extension of the current one, in which information is given welldefined meaning, better enabling computers and people to work in cooperation. (Berners-Lee, Handler und Lassila, 2001)

In anderen Worten:

Das Semantic Web ist die Erweiterung des World Wide Webs als eines Netzes von Dokumenten zu einem Netz von Daten, von maschinenlesbaren, einzelnen Aussagen. Durch maschinelles logisches Schlussfolgern soll aus diesen Daten zusätzliches Wissen generiert werden können. (A. Pohl, 2011, S. 2)

Für weitere einführende Beschreibungen sei an dieser Stelle verwiesen auf den genannten Artikel von Berners-Lee et al., die einleitenden Kapitel von Allemang und Hendler (2008) sowie auf Campbell und MacNeill (2010).

Basierend auf den vorstehenden Erläuterungen lässt sich die Bedeutung des Begriffs «Linked Data» leicht ableiten. Heath und Bizer (2011, Kapitel 2) formulieren wie folgt:

The term Linked Data refers to a set of best practices for publishing and interlinking structured data on the Web. These best practices were introduced by Tim Berners-Lee in his Web architecture note Linked Data and have become known as the Linked Data principles. These principles are the following:

1. Use URIs as names for things.
2. Use HTTP URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

Ein weiterer, in der Diskussion oft genannter Begriff ist «Open Data». Unter diesem Begriff wird in der Regel die Erwartung diskutiert, dass bereitgestellte Daten (nicht Dienste) möglichst frei von Restriktionen zur weiteren Nachnutzung sein sollten. Eine Einführung in dieses Themengebiet bietet A. Pohl (2010). Die Verknüpfung der beiden Begriffe zu «Linked Open Data» fügt also eine lizenzrechtliche Dimension zum Linked Data-Konzept hinzu und erfordert in der praktischen Konsequenz eine Möglich-

keit, diese Informationen in ebenfalls maschinenlesbarer Art mit den eigentlichen Daten zu assoziieren.

RDF oder das «Resource Description Framework» schließlich ist ein Standard des W3C<sup>74</sup>, in dem ein Datenmodell für die Repräsentation von Linked Data definiert wird. Das Datenmodell basiert auf Drei-Elementigen Aussagen (sog. «Tripeln»), bestehend aus Subjekt, Prädikat (engl. hier «Property») und Objekt. Dies erlaubt es, auf semantische Art ein Objekt, Konzept oder Wert mit einem anderen in Beziehung zu setzen (vgl. Semantisches Web). Einige Beispiele für mögliche Aussagen können der Tabelle 2 entnommen werden.

Subjekt	Prädikat	Objekt
ISBN:0-12345-678-9	hat Abstract	In diesem Buch werden die Prinzipien ...
ISBN:0-12345-678-9	hat Umschlagbild	<a href="http://amazon.com/cover/0123456789.png">http://amazon.com/cover/0123456789.png</a>
HBZ:HT1230000019	ist alternative ID für	ISBN:0-12345-678-9

**Tabelle 2** (Fiktive) Beispiele für Aussagen in Tripel-Form

Ein Triple Store ist eine für die Speicherung von Tripeln – also nach RDF-Modell gestalteten Aussagen – ausgelegte Software, die mit SPARQL<sup>75</sup> (eine rekursive Abkürzung für «SPARQL Protocol And RDF Query Language») i. d. R. auch eine spezielle Sprache für den gezielten Abruf von Tripeln unterstützt.

Wie eingangs erwähnt geht es darum, Verbindungen mit bestimmten Bedeutungen zwischen Subjekt und Objekt herzustellen. Daher wird für die Prädikate (oder «Eigenschaften») i. d. R. ein klar definiertes und möglichst verbreitetes Vokabular verwendet, dessen einzelne Einträge über weltweit eindeutige Identifikatoren (URIs) verfügen und dadurch maschinell ausgewertet werden können. Diese Vokabularien werden im vorliegenden Kontext als «Ontologien» bezeichnet. Bevor die gewünschten Aussagen also getroffen werden können, müssen diese zunächst definiert werden: welche Aussage sie repräsentieren sowie ggf. über welche Arten (Klassen) von Subjekten sie getroffen werden und welche möglichen Werte sie annehmen können. Anschließend weist die definierende Stelle jeder Aussage eine eindeutige URI möglichst aus einem durch sie kontrollierten Namensraum zu. Bereits erwähnt wurde, dass es sich dabei möglichst um HTTP URIs handeln soll. Ein Beispiel für eine solche URI könnte so aussehen: <http://caesar.org/enrichments/hasCover>. HTTP URIs können in gewissem Rahmen menschenlesbar und selbsterklärend sein, sind aber trotzdem primär für eine maschinelle Verarbeitung vorgesehen und geeignet.

<sup>74</sup> Siehe RDF Primer, 2004.

<sup>75</sup> Siehe <http://www.w3.org/standards/techs/sparql>.

---

Objekte in Tripeln können entweder Literale (beliebige Zeichenketten) sein, die eine bestimmte Information enthalten oder URIs, die andere Ressourcen bezeichnen. Subjekte können nur URIs – als Bezeichner beliebiger Ressourcen – sein.

Ein entscheidender Punkt ist, dass es – sobald für eine beliebige Ressource, egal ob digital, physisch oder rein konzeptuell – eine URI festgelegt wurde<sup>76</sup> möglich ist, jederzeit jede beliebige Aussage über die Ressource zu machen. Allemang und Hendler (2008, S. 7) betonen diesen Umstand mehrfach und verwenden den Ausdruck «AAA-Slogan» dafür: «Anyone can say Anything about Any topic».<sup>77</sup> Durch die Verknüpfung von Aussagen ist es dabei möglich, zu neuen, nicht explizit festgelegten Erkenntnissen zu gelangen.

## 9.2 RDF Triple Store für die Anreicherungsdatenbank

Welche Vorteile oder Nachteile hätte nun die Verwendung eines Triple Store für den Anreicherungsdienst? Ein eher psychologischer Faktor ist die weit verbreitete und zunehmende Popularität von RDF und Linked Data, die eine solche Lösung strategisch und politisch attraktiv erscheinen lässt.

Zu den großen Vorteilen aus Sicht der Datenhaltung gehört, dass es völlig unproblematisch ist, weitere Informationen (in Form von Tripeln) in den Datenbestand zu integrieren, ohne das dabei interne Strukturen zerstört oder verändert werden müssten. Vorhandene Abfragen funktionieren damit einfach weiter bzw. können, wenn sie hinreichen generisch formuliert wurden, die neuen Informationen unmittelbar mit erfassen.

Viele Datenbestände, die als Rohdaten für die Anreicherungsdatenbank in Frage kommen, sind in diesem Format verfügbar und können daher mit entsprechender Software gut verarbeitet werden. Auch können Datenexporte aus der Anreicherungsdatenbank in diesem Format problemlos weitergegeben werden. Heath und Bizer (2011, Abschnitt 2.4.1) schreiben dazu explizit:

The data model is designed for the integrated representation of information that originates from multiple sources, is heterogeneously structured, and is represented using different schemata. RDF aims at being employed as a lingua franca, capable of moderating between other data models that are used on the Web.

Dies entspricht exakt den Rahmenbedingungen, die in den vorhergehenden Kapiteln dargelegt wurden.

Ein Triple Store mit Unterstützung für die SPARQL-Abfragesprache könnte auch direkt – ohne die zusätzliche kapselnde Schicht einer Pro-

---

<sup>76</sup> Dieser Vorgang wird oft als «prägen» einer URI bezeichnet.

<sup>77</sup> Als gleichwertiger Ausdruck wird auch die Wendung «Open World Assumption» gebraucht (vgl. „Named Graphs“, 2005, S. 250).

---

grammierschnittstelle (API) – freigegeben werden. So könnten die gesammelten Daten für vielfältiger Zwecke flexibel durch andere Parteien genutzt werden – vorausgesetzt, es stehen der Anreicherungsdatenbank ausreichende Verarbeitungskapazitäten zur Verfügung.

Um Anreicherungen für bibliographische Ressourcen ausliefern zu können, muss eine Verbindung zwischen beiden Ressourcen hergestellt werden. Bibliographische Ressourcen werden, wie zuvor dargelegt, grundsätzlich durch Identifikatoren repräsentiert. Um Aussagen über Identifikatoren (= Subjekte) in Tripel-Form treffen zu können, müssen URIs für diese Identifikatoren geprägt werden und nicht für alle Identifikatoren wurde dies bereits getan. Zwar ist dies auf der technischen Ebene simpel. Die Erzeugung «guter» URI erfordert jedoch einige Überlegungen, wie etwa im Bericht *Designing URI Sets for the UK Public Sector* (2009) deutlich wird.

Die im Abschnitt 9.1 zitierten Prinzipien verlangen zudem, URIs zu verwenden, die langfristig aufgerufen werden können und dann sinnvolle Informationen über die durch die URIs identifizierten Ressourcen liefern (vgl. auch das Bewertungsschema für publizierte Datenbestände bei Berners-Lee, 2006). Die Prinzipien sind tatsächlich als wesentliche Anforderungen zu verstehen und sollten keinesfalls auf die leichte Schulter genommen werden. Für die vorliegenden Zwecke kann es allerdings sinnvoll sein, URN statt URL als Identifikatoren zu verwenden und damit den bei URL implizierten Aufwand zur tatsächlichen Bereitstellung von Informationen unter der entsprechenden Adresse zu vermeiden. Beispielsweise für ISBN («urn:ISBN:006251587X») ist dies bereits etablierte Praxis (vgl. Listing 1 bzw. Hakala und Walravens, 2001).<sup>78</sup>

Einmal öffentlich zugänglich gemacht, ist die Verwendung der URIs auch nicht mehr kontrollierbar, so dass Änderungen zu Irritationen und Abstimmungsproblemen führen. Die Notwendigkeit zur Prägung von URIs sowie der Bereitstellung einer Infrastruktur zur langfristigen Auflösbarkeit stellt somit durch den damit verbundenen Aufwand einen indirekten Nachteil der Verwendung eines Triple Stores dar.

Die Unterstützung von Freitext-Abfragen über (proprietäre) Syntax-Erweiterungen von SPARQL ist noch nicht allgemein verbreitet und/oder ausgereift in Triple Stores. Sofern also beispielsweise vollständige Abstracts, Rezensionen o.ä. in der Anreicherungsdatenbank hinterlegt und dort durchsucht werden sollen, ist dies möglicherweise nicht oder nur mit Abstrichen möglich. Die Performance einer Lösung auf Basis einer Suchmaschinensoftware ist nicht zu erwarten. Besteht dieser Bedarf, so schränkt er entweder die Auswahl der Triple Store-Software auf wenige Möglichkeiten ein oder erfordert die Implementierung einer zusätzlichen Komponente. So könnte ein erster Schritt über eine Freitextsuche in einer Suchmaschine ei-

---

<sup>78</sup> In diesem Kontext sollte die weitere Entwicklung des «Actionable ISBN»-Ansatzes der deutschen ISBN-Agentur beobachtet werden (vgl. „Actionable ISBN Application Introduced in Germany“ 2011).

---

nes Liste an Identifikatoren (URIs) ermitteln, über die dann im zweiten Schritt weitere Daten aus dem Triple Store abgerufen werden.

### 9.3 Repräsentation der Identität von Ressourcen in RDF

Eines der zentralen Probleme für die Datenhaltung im Anreicherungsdienst ist die in Kapitel 8 diskutierte Abbildung von Identitäten bzw. die Gruppierung von Identifikatoren.

RDF ist, wie erwähnt, als universelles Datenmodell zum Austausch von Daten zwischen verschiedenen Produzenten gedacht. Jeder Produzent von Daten muss URIs für die Subjekte prägen, über die er Aussagen festhalten will, aber es gibt keine einzelne globale Einrichtung, die die Vergabe von URIs koordiniert. Es ist daher nicht ungewöhnlich, dass mehrere URIs für die gleiche Ressource geprägt werden.

The explosion in the number of information sources being exposed as RDF has also led to an explosion in the number of URIs used to identify different entities. It is often the case that data in different repositories will hold information regarding identical entities. [...] The multiplicity of URIs leads to the problem of coreference, where different URIs are used to describe the same entity. (Jaffri, Glaser und Millard, 2007, S. 2)

Um die Aussagen über eine Ressource zusammenführen zu können, müssten entweder die Duplikate entfernt und durch eine einzige URI ersetzt werden – was aufgrund der verteilten Datenhaltung im WWW und des notwendigen Koordinationsaufwands unmöglich ist, oder es muss eine Möglichkeit geben, verschiedene URIs als gleichbedeutend zu kennzeichnen. Der allgemein akzeptierte Ansatz für letzteres ist die Verwendung des Prädikats «sameAs» aus der «Web Ontology Language»<sup>79</sup> (OWL).<sup>80</sup>

Allerdings ist owl:sameAs strikt in der damit verbundenen Aussage, dass durch ein Tripel «<RessourceA> owl:sameAs <RessourceB>» beide Ressourcen als identisch und vollständig austauschbar gekennzeichnet werden (vgl. When owl:sameAs isn't the Same, 2010, S. 2). Für die Zwecke der Anreicherungsdatenbank zur Verknüpfung der Identifikatoren der bibliographischen Ressourcen kann owl:sameAs deshalb nicht verwendet werden. Zum einen bezeichnen die Identifikatoren unterschiedliche «Dinge». Eine Erstkatalogisierungs-ID (EKI) ist ein Bezeichner für einen Katalogisierungsdatensatz. Eine ISBN dagegen bezeichnet in etwa eine FRBR-Manifestation. Zum anderen sollen mehrere ISBN im Sinne eines FRBR-Werkes verknüpft werden, aber diese ISBN bezeichnen keinesfalls identische Ressourcen.

---

<sup>79</sup> Vgl. <http://www.w3.org/TR/owl2-overview/>.

<sup>80</sup> Für eine detailliertere Erläuterung Heath und Bizer (vgl. 2011, Abschnitt 2.5.2).



---

Jaffri, Glaser und Millard (2007) führen daher das Konzept der «Bundles» (Bündel) ein, die eine Menge von URIs beinhalten, die in einem definierten Kontext als austauschbar angesehen werden können und verwenden den Begriff «Koreferenzen» für die Beschreibung dieser Situation. Dieses Konzept entspricht offensichtlich dem Bedarf im Kontext der Anreicherungsdatenbank.

Konfrontiert mit einem ähnlichen Anwendungsfall wurden im Culturegraph-Projekt drei in Frage kommende Ontologien zum Ausdruck der Koreferenz identifiziert (vgl. Koreferenzen im Semantic Web 2011): die coref-Ontologie<sup>81</sup> von Jaffri, Glaser und Millard, die Bundle-Ontologie<sup>82</sup> von Ben O’Steen (vgl. O’Steen, 2010) sowie die Similarity-Ontologie<sup>83</sup> von Jacobson, Raimond und Gängler (2010). Bei Culturegraph scheint es derzeit eine Präferenz zur Verwendung der Bundle-Ontologie zu geben, weil diese im Vergleich zur coref-Ontologie zusätzlich die Angabe der Begründung (formale Herleitung) für die Koreferenz erlaubt (Koreferenzen im Semantic Web 2011). Sie ist umgekehrt weniger komplex und umfangreich als die Similarity-Ontologie.

Im Kontext der Anreicherungsdatenbank ist die Begründung für die Feststellung der Koreferenz nicht von Interesse, da sich daraus keine praktischen Konsequenzen für die Funktionalität ergeben. Zudem ist nicht absehbar, inwieweit diese Information überhaupt aus den auswertbaren Rohdaten extrahiert werden könnte. Es macht daher Sinn, an dieser Stelle die simpelste ausreichende Lösung – die coref-Ontologie – zu wählen.

## 9.4 Anreicherungsressourcen in RDF

Ziel des in dieser Arbeit entwickelten Dienstes bzw. von dessen Datenbank ist die Bereitstellung von Informationen über die an einer beliebigen Stelle verzeichneten bibliographischen Ressourcen. Wie bereits im Abschnitt 9.1 beschrieben, müssen mögliche Aussagen a priori festgelegt werden, damit sie (auch in verteilten Umgebungen) konsistent verwendet werden können.<sup>84</sup>

Inzwischen existiert eine Vielzahl teils überlappender, teils divergierender Ontologien für verschiedenste Einsatzgebiete.<sup>85</sup> Dies macht es in der Tat bereits schwierig, die für einen bestimmten Zweck am besten geeignete Ontologie zu identifizieren oder überhaupt herauszufinden, ob eine solche Ontologie bereits existiert (vgl. Heath und Bizer, 2011, Abschnitt 4.4.5). Hinzu

---

<sup>81</sup> <http://www.rkbexplorer.com/ontologies/coref#>.

<sup>82</sup> <http://benosteen.com/bundle.rdf>.

<sup>83</sup> <http://kakapo.dcs.qmul.ac.uk/ontology/musim/0.2/musim.owl>.

<sup>84</sup> Der Erfahrung des Autors nach wäre es allerdings ein Fortschritt, wenn Verweise auf Anreicherungsressourcen überhaupt durchgängig qualifiziert würden, denn dies ist bislang in vielen Katalogen keineswegs selbstverständlich, so dass Benutzer oft genug erst nach Aufruf eines Links erkennen können, welche Art von Information dort vorhanden ist.

<sup>85</sup> Vgl. etwa <http://schemapedia.com/>, <http://vocab.deri.ie/> oder <http://metadataregistry.org/>.

---

kommt, dass in vielen Fällen eine einzelne Ontologie nicht ausreicht, um alle gewünschten Aussagen formulieren zu können.

#### 9.4.1 Beispiel

Als Beispiel sei an dieser Stelle auf die RDF-Repräsentation eines durch den «RDF Book Mashup» Dienst<sup>86</sup> generierten Datensatzes<sup>87</sup> in Listing 1 verwiesen<sup>88</sup>.

Die Einträge in den Zeilen 3 bis 8 bezeichnen die verwendeten Ontologien, aus denen die Prädikate zur Formulierung von Aussagen über mehrere Ressourcen / Subjekte entnommen wurden. In den Zeilen 10, 26, 35, 38 und 41 stehen jeweils die URIs (Identifikatoren) der Subjekte, über die die Aussagen in den folgenden Zeilen getroffen werden.

Die URI in Zeile 26 ist (absichtlich) identisch mit der URL unter der das vorliegende Dokument abgelegt wurde. Die folgenden Zeilen 27 bis 33 enthalten Informationen über das Dokument – von wem es erstellt wurde (Zeile 32), über welche Ressource es primär Aussagen enthält (Zeile 33) oder welche Lizenzbedingungen mit den Informationen innerhalb des Dokuments verknüpft sind (Zeile 27 bis 29).

Zeile 10 enthält eine andere URI, die folgenden Zeilen 11 bis 24 enthalten also Aussagen über dieses andere Subjekt. Zeile 21 informiert uns beispielsweise, dass das Subjekt vom Typ «Book» nach dem Verständnis des «Simple Commerce Vocabulary» ist.

---

<sup>86</sup> <http://www4.wiwiss.fu-berlin.de/bizer/bookmashup/>; vgl. zur Erläuterung Bizer, Cyganiak und Gauß (2007).

<sup>87</sup> <http://www4.wiwiss.fu-berlin.de/bookmashup/doc/books/006251587X>.

<sup>88</sup> Aufgrund der dynamischen Generierung der Datensätze kann das Ergebnis eines erneuten Aufrufs der URL von der Auflistung abweichen. Der im Listing gezeigte Datensatz wurde zudem aus Platzgründen gekürzt und wird statt in der originalen Serialisierung in RDF/XML in der sogenannten Turtle-Syntax (vgl. <http://www.w3.org/TeamSubmission/turtle/>) wiedergegeben. Für die Darlegung der Überlegungen im Text ist die präzise Wiedergabe jedoch unerheblich, da es um das grundlegende Prinzip geht.

```

01 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
02 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
03 @prefix dc: <http://purl.org/dc/elements/1.1/> .
04 @prefix dcterms: <http://purl.org/dc/terms/> .
05 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
06 @prefix rev: <http://purl.org/stuff/rev#> .
07 @prefix scom:
08   <http://sites.wiwiss.fu-berlin.de/suhl/bizer/bookmashup/simpleCommerceVocab01.rdf#> .
09
10 <http://www4.wiwiss.fu-berlin.de/bookmashup/books/006251587X>
11   dc:creator "Tim Berners-Lee" ;
12   dc:date "2000-11-01" ;
13   dc:format "Paperback" ;
14   dc:identifier <urn:ISBN:006251587X> ;
15   dc:publisher "Harper Paperbacks" ;
16   dc:title "Weaving the Web" ;
17   rev:hasReview
18     <http://www4.wiwiss.fu-berlin.de/bookmashup/reviews/006251587X_EditorialReview1> ;
19   scom:hasOffer
20     <http://www4.wiwiss.fu-berlin.de/bookmashup/offers/006251587XamazonOffer> ;
21   a scom:Book ;
22   rdfs:label "Weaving the Web" ;
23   foaf:depiction <http://ecx.images-amazon.com/images/I/5188BQ7668L.jpg> ;
24   foaf:thumbnail <http://ecx.images-amazon.com/images/I/5188BQ7668L._SL75_.jpg> .
25
26 <http://www4.wiwiss.fu-berlin.de/bookmashup/doc/books/006251587X>
27   dcterms:license
28     <http://www.amazon.com/AWS-License-home-page-Money/b/ref=...>,
29     <http://www.google.com/terms_of_service.html> ;
30   a foaf:Document ;
31   rdfs:label "RDF document about the book: Weaving the Web" ;
32   foaf:maker <http://www4.wiwiss.fu-berlin.de/is-group/resource/projects/Project10> ;
33   foaf:primaryTopic <http://www4.wiwiss.fu-berlin.de/bookmashup/books/006251587X> .
34
35 <http://www4.wiwiss.fu-berlin.de/bookmashup/offers/006251587XamazonOffer>
36   rdfs:label "Offer for the book with the ISBN: 006251587X" .
37
38 <http://www4.wiwiss.fu-berlin.de/bookmashup/reviews/006251587X_EditorialReview1>
39   rdfs:label "Review number 1 about: Weaving the Web" .
40
41 <http://www4.wiwiss.fu-berlin.de/is-group/resource/projects/Project10>
42   rdfs:label "RDF Book Mashup" .

```

### Listing 1 Gekürzter «RDF Book Mashup» Datensatz

Die Zeilen 17/18 bzw. 23 und 24 enthalten Aussagen, die so auch im Rahmen des Anreicherungsdienstes vorkommen könnten. Das «RDF Review Vocabulary»<sup>89</sup> mit der Namensraum-URI <http://purl.org/stuff/rev#> (vgl. Zeile 6) definiert offenbar ein Prädikat `hasReview`, das hier verwendet wird um das Subjekt mit der URI [http://www4.wiwiss.fu-](http://www4.wiwiss.fu-berlin.de/bookmashup/reviews/006251587X_EditorialReview1)

<sup>89</sup> Vgl. <http://vocab.org/review/terms.html>.

---

berlin.de/bookmashup/books/006251587X<sup>90</sup> mit dem Objekt zu verbinden, dessen URI [http://www4.wiwiss.fu-berlin.de/bookmashup/reviews/006251587X\\_EditorialReview1](http://www4.wiwiss.fu-berlin.de/bookmashup/reviews/006251587X_EditorialReview1)<sup>91</sup> ist.

```
01 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
02 @prefix rev: <http://purl.org/stuff/rev#> .
03
04 <http://www4.wiwiss.fu-berlin.de/bookmashup/reviews/006251587X_EditorialReview1>
05   rev:text "'<P>Named one of the greatest minds of the 20th century by <I>Time</I>, Tim '
06   Berners-Lee is responsible for one of that century's most important advancements: the '
07   world wide web. Now, this low-profile genius-who never personally profitted from his '
08   invention -offers a compelling protrait of his invention. He reveals the Web's '
09   origins and the creation of the now ubiquitous http and www acronyms and shares his '
10   views on such critical issues as censorship, privacy, the increasing power of '
11   software companies , and the need to find the ideal balance between commercial and '
12   social forces. He offers insights into the true nature of the Web, showing readers '
13   how to use it to its fullest advantage. And he presents his own plan for the Web's '
14   future, calling for the active support and participation of programmers, computer '
15   manufacturers, and social organizations to manage and maintain this valuable resource '
16   so that it can remain a powerful force for social change and an outlet for individual '
17   creativity.</P>'" ;
18   rdf:type rev:Review .
```

#### Listing 2 Externe verknüpfte Informationen zu Listing 1

Für alle vorstehenden Ausführungen gilt, dass die konkrete Bedeutung dieses Tripels sich erst durch die Analyse der Klartext-Beschreibungen der verwendeten Vokabularen erschließt. Da es sich bei der URI des Objektes um eine dereferenzierbare HTTP URI handelt (vgl. die Prinzipien in Abschnitt 9.1), kann man die URI als URL behandeln und darüber ein weiteres RDF-Dokument abrufen, das u. a. die in Listing 2 gezeigten Informationen enthält.

Die Aussage in den Zeilen 23 und 24 von Listing 1 sind strukturell identisch. Allerdings verweisen die Objekt-URI hier nicht auf weitere RDF-Dokumente, sondern direkt auf Grafikdateien.

Es sei angemerkt, dass es keinen technischen Grund gibt, der gegen eine Integration der in Listing 2 enthaltenen Tripel in das in Listing 1 gezeigte Dokument spricht. Die Vergabe einer eigenen URI ist davon ebenfalls völlig unberührt.

Bemerkenswert ist in Listing 1 des weiteren die Zeile 14. Dort findet sich mit urn:ISBN:006251587X ein Beispiel für eine URI, die nicht als HTTP URL gestaltet ist, sondern als URN gemäß RFC 3187 (vgl. Hakala und Wal-

---

<sup>90</sup> Die Ressource, die hier als Subjekt verwendet wird, ist kein elektronisches Objekt, das unter dieser URL existiert, sondern die Publikation von Berners-Lee. Gemäß dem Konzept der «303 URIs» wird daher ein HTTP-Aufruf der URL <http://www4.wiwiss.fu-berlin.de/bookmashup/books/006251587X> auf die URL <http://www4.wiwiss.fu-berlin.de/bookmashup/doc/books/006251587X> umgeleitet, unter der ein Dokument mit Informationen über die erstgenannte URI ausgeliefert wird. Für eine Erläuterung der Hintergründe vgl. Cool URIs for the Semantic Web (2008).

<sup>91</sup> Analog zur Subjekt-URI erfolgt hier eine Umleitung auf [http://www4.wiwiss.fu-berlin.de/bookmashup/doc/reviews/006251587X\\_EditorialReview1](http://www4.wiwiss.fu-berlin.de/bookmashup/doc/reviews/006251587X_EditorialReview1).

---

ravens, 2001). Diese URI ist nicht direkt dereferenzierbar und damit insbesondere auch für Menschen weniger «zugänglich», erfüllt aber trotzdem den Zweck der globalen Eindeutigkeit. Auch der in dieser Arbeit entwickelte Anreicherungsdienst könnte sich URN für die Zuordnung von Anreicherungen zu bibliographischen Ressourcen zunutze machen.

#### 9.4.2 Formale Inkonsistenzen

Um Ontologien zu definieren, reicht eine Freitext-Beschreibung ihrer Elemente nicht aus, weil diese zu viele Unklarheiten zuließe. Stattdessen wurde u.a. vom W3C eine Reihe aufeinander aufbauender und selbst wieder mittels des RDF-Datenmodells formalisierbarer Vokabularien geschaffen, die es erlauben, die Eigenschaften und Relationen von abstrakten Ressourcen zu beschreiben. Die Basis aller Arbeiten auf diesem Gebiet ist RDF Schema (Brickley und Guha, 2004). Mittels RDF Schema kann unter anderem festgelegt werden, auf welche Arten («Klassen») von Subjekten ein Prädikat angewendet werden kann und welche Arten von Aussagen es treffen kann (i.e., welchen Klassen die Objekte angehören müssen).

Als Subjekte haben wir im Verlauf der Arbeit «bibliographische Ressourcen» festgelegt, die durch bestimmte Identifikatoren repräsentiert werden. In Abschnitt 9.3 wurde festgestellt, dass bestimmte Identifikatoren nicht als vollkommen austauschbar angesehen werden können, weil sie in letzter Konsequenz nicht auf die gleiche Ressource verweisen: eine Datensatz-ID in einem Katalog identifiziert eben den Datensatz, während eine ISBN i. d. R. eine FRBR-Manifestation identifiziert. Folgt man diesem Gedanken theoretisch/puristisch bis zum Ende, gelangt man zu der Erkenntnis, dass eine Aussage (ein Prädikat) «hat ein Umschlagbild» nicht über ein Subjekt mit einer URI auf Basis eines Datensatz-Identifikators gemacht werden kann: ein Datensatz hat keinen Umschlag.

Folgerichtig müsste beispielsweise für Manifestation ohne ISBN zunächst eine URI «erfunden» werden. Danach können über das durch diese URI repräsentierte Subjekt dann Aussagen getroffen werden – zum Beispiel, dass es durch einen bestimmten Datensatz mit einer bestimmten Datensatz-ID beschrieben wird oder dass es eine Umschlagabbildung zu dem Subjekt gibt.

Zumindest im Kontext des zu entwickelnden Anreicherungsdienstes entsteht aus diesem theoretisch saubereren Ansatz jedoch keinerlei Mehrwert. Eine Möglichkeit wäre, die Fragestellung mangels praktischer Konsequenzen schlicht zu ignorieren. Eine andere Möglichkeit wäre, die generierten und mit URI versehenen coref-Bundles als Platzhalter für ein gedankliches Konstrukt einer «bibliographischen Werk-Ressource» zu definieren und die Aussagen («hat ein Umschlagbild») jeweils mit der URI des Bundles als Subjekt zu treffen. Damit ginge jedoch ein gewisser Informationsverlust einher: Sollen coref-Bundles idealtypisch auf der Ebene von FRBR-Werken liegen, so kann beispielsweise bei der Bindung eines Umschlagbildes an eine ISBN und einer Abfrage des Anreicherungsdienstes mit genau dieser

---

ISBN das entsprechende in der Anzeige bevorzugt werden. Ist das Bild dagegen dem Bundle zugeordnet, ist diese Bevorzugung nicht mehr möglich.

Im Endeffekt scheint – in Abwägung von theoretischen Ansprüchen, praktischen Anforderungen und Arbeitsaufwand – der sinnvollste Ansatz darin zu bestehen, den Gedanken zu ignorieren und Datensatz-Identifikatoren als Stellvertreter für die Ressourcen an sich zu akzeptieren.

#### 9.4.3 Metadaten über einzelne Anreicherungsressourcen

Bislang sind für die Datensammlung des Anreicherungsdienstes logisch gesehen nur zwei Arten von Bezügen zwischen Ressourcen vorgesehen: zum einen die Verknüpfung zwischen Identifikatoren (in Form von URI) zum Ausdruck ihrer Koreferenz und zum anderen die Verknüpfung zwischen Anreicherungsressourcen (in Form von Literalen oder URI) und den vorgenannten Identifikatoren.

Die Anreicherungsressourcen verfügen jedoch selbst ebenfalls wieder über Eigenschaften, deren Verzeichnung und Auswertung für die Optimierung der Funktionalität des Anreicherungsdienstes genutzt werden könnte. Welche Eigenschaften das sind, ist natürlich abhängig vom Typ der Anreicherungsressource. Als Beispiele sollen hier Größe<sup>92</sup> von Umschlagabbildungen und die Sprache einer Inhaltsangabe dienen. Sind die vorgenannten Informationen verfügbar, können Anwender des Anreicherungsdienstes davon Gebrauch machen, indem sie etwa für eine Thumbnail-Vorschau gezielt kleine Bilder heranziehen oder ihren Benutzern bevorzugt Inhaltsangaben in einer bestimmten Sprache anbieten.

Die Aufnahme solcher Metadaten über die Anreicherungsressourcen erfordert «lediglich» die Auswahl passender Prädikate und das Einfügen der entstehenden Tripel in die Datenbank.

Ob und wie diese Informationen ausgewertet werden, ist eine Frage, die im Zuge der Gestaltung der Anwendung bzw. der Schnittstellen beantwortet werden muss (vgl. dazu Kapitel 10).

#### 9.4.4 Metadaten über Teilbestände der Anreicherungsressourcen

Die gesammelten und in der Datenbank (bzw. konkreter: dem Triple Store) zusammengeführten Anreicherungsressourcen entstammen aus diversen Quellen. Es ist davon auszugehen, dass nicht alle Lieferanten ihre Daten vollständig ohne jegliche Vorbehalte bereitstellen werden (vgl. Kapitel 7). Die mit der Nutzung verbundenen Auflagen sind potentiell beliebig: während einem Lieferanten die Nennung der Datenquelle genügt, kann ein anderer verlangen, jede von ihm bereitgestellte Anreicherungsressource mit einer bestimmten Webseite zu verlinken.

Solange die Anforderung nicht spezifisch für jede einzelne Anreicherungsressource ausfallen oder nicht maschinell daraus oder aus den Identifi-

---

<sup>92</sup> Angabe von Breite und Höhe in Pixeln.

---

katoren, mit denen sie verknüpft sind, abgeleitet werden können<sup>93</sup>, macht es Sinn, die für die Erfüllung der Auflagen notwendigen Informationen einmalig abzulegen und mit dem gesamten Datenpool des entsprechenden Lieferanten als Einheit zu verlinken.

Die Möglichkeit zur retrospektiven Identifizierung eines bestimmten Teilbestands im gesamten Datenbestand des Anreicherungsdienstes ist zusätzlich noch aus zwei weiteren Gründen sinnvoll. Zum einen erlaubt es diese Fähigkeit, gezielt Daten zu löschen (z. B. weil sich die Nutzungsbedingungen des Lieferanten geändert haben), ohne dass dafür die Datenbank von Grund auf neu aufgebaut werden muss. Zum anderen können bei Bedarf Einschätzungen zur Qualität und Zuverlässigkeit der Datenherkunft in die konkrete Anzeige der Daten einbezogen werden.

Die Bereitstellung von Metadaten über RDF-Dokumente ist nicht besonders komplex. Eine Erläuterung zur Vorgehensweise liefern Heath und Bizer (2011, Abschnitt 4.3). Die Frage, wie genau solche Metadaten (Lizenzen, Nutzungsbedingungen, Provenienz usw.) in Bezug auf RDF-Graphen (Mengen von Tripeln) als Teilbestände innerhalb einer größeren Datenmenge mit den Mitteln von RDF ausgedrückt werden könnten, wurde über mehrere Jahre diskutiert. Im Jahr 2005 wurde mit dem Konzept der «Benannten Graphen» («Named Graphs») jenes Modell vorgestellt („Named Graphs“, 2005, vgl.), das sich aktuell durchzusetzen scheint (Bizer, 2010, Folie 2). Es wird zudem in SPARQL (der Standard-Abfragesprache für Triple Stores) verwendet (vgl. ebenfalls Bizer, 2010, Folie 2).

Welche Vokabularen für die Formulierung der Metadaten über die Teilbestände verwendet werden, hängt also unter anderem davon ab, welche Nutzungsbedingungen durch die jeweiligen Lieferanten für sie festgelegt werden. Heath (2010, Abschnitte 4.3 und 4.4) stellen einige in Frage kommende Vokabularen vor. Sinnvollerweise sollten die Ausgangsdaten bewahrt werden. Zur Vereinfachung der Nutzung der Daten – speziell im Hinblick auf eine Abfrage per API oder SPARQL durch Dritte statt durch eine vom Betreiber des Anreicherungsdienstes selbst entwickelte Client-Software – könnte es allerdings sinnvoll sein, die verschiedenen Ansprüche zusätzlich in ein über alle Teilbestände kohärentes Vokabular zu überführen. Nutzungsbedingungen sind ernst zu nehmen und die Vereinfachung ihrer Beschreibung verringert das Risiko ihrer Verletzung erheblich.

Andere Metadaten können z. B. der vorgeschriebene Text für eine Herkunftsangabe oder ein «Template» für die Generierung von Links zum Lieferanten.

#### 9.4.5 Erkenntnisse und Folgerungen

Die Vielzahl verfügbarer Vokabularen und die Tatsache, dass es keine einzelne zentrale Stelle gibt, an der sie verzeichnet sind, führen dazu, dass im-

---

<sup>93</sup> Beispielsweise, indem der jeweils verlangte Link aus einer Basis-URL und der ISBN der jeweiligen bibliographischen Ressource zusammengefügt wird.

---

mer dann, wenn zum ersten Mal ein Objekt eines neuen Typs von Anreicherungsressourcen in die Datenbank aufgenommen werden soll, erkundet werden muss, ob es bereits eine Ontologie gibt, die es erlaubt, den Bezug zwischen diesem Objekt und einer bibliographischen Ressource herzustellen. Ist ein entsprechendes Prädikat verfügbar, sollte es verwendet werden. Kommen mehrere Prädikate in Frage, so spricht aufgrund des RDF Datenmodells nichts dagegen, die gewünschte Aussagen redundant mit allen Prädikaten in der Datenbank abzulegen. Andererseits wird die Datenbank dadurch aufgebläht, was zu Performance-Einbußen führen kann. Wenn ein Prädikat erkennbar weiter verbreitet ist als die Alternativen, macht es daher Sinn, sich auf dieses Prädikat zu beschränken.<sup>94</sup> Geprüft werden muss zudem, ob mit den Prädikaten Aussagen über die kombinierbaren Subjekte und Objekte verknüpft sind, die mit den vorliegenden Rahmenbedingungen unvereinbar sind und die Verwendung eines Prädikats daher ausschließen.

Die Definition und Verwendung eigener Prädikate wird innerhalb der RDF «Gemeinde» grundsätzlich als Ultima Ratio betrachtet. Damit Tripel von anderen Parteien gezielt nachgenutzt werden können, müssen die verwendeten Ontologie bekannt sein. Die Information über und das Erlernen von neuen Ontologien kostet alle beteiligten Parteien Ressourcen. Ebenso steigt der konkrete Arbeitsaufwand bei der Zusammenführung von verteilt erstellten Datenbeständen, da dabei alternative Ausdrucksformen gegeneinander abgeglichen werden müssen.

Die vorstehenden Bedenken könnten allerdings zurückgestellt werden, wenn jegliche externe Nutzung der Datenbank (zunächst) nur über eine API abliefe, die die internen Datenstrukturen kapseln könnte oder deren Nutzung externen Anwender sogar untersagt würde. Ein solches Vorgehen brächte zumindest vorübergehend beträchtliche Freiheiten in der Auswahl des RDF-Vokabulars und bei der Prägung von URI und würde somit den Aufwand zur Ermittlung und Auswahl geeigneter Vokabularien und Methoden verringern. Langfristig ist selbstverständlich die Verwendung etablierter Praktiken und Standards anzustreben, da dies unabhängig von einem Datenaustausch zumindest dem intellektuellen Austausch mit der Linked Data Entwicklungsgemeinschaft förderlich ist.

Im Anhang IV finden sich einige knappe Beispiele, wie coref-Bundles und benannte Graphen mit Anreicherungsressourcen tatsächlich aussehen. Dazu werden zwei minimale SPARQL-Abfragen gezeigt und schließlich ein Perl-Skript, das die RDF-Daten in einen Triple Store importiert und mit den gezeigten SPARQL-Abfragen die Informationen wieder ausliest.

---

<sup>94</sup> Die inhärente Aussagenlogik in RDF erlaubt es, aus vorhandenen Tripeln vielfältige Schlüsse zu ziehen. Für weitere Erläuterungen zum Thema «Inferencing» sei an dieser Stelle auf Kapitel 5 bei Allemang und Hendler (2008) verwiesen.



---

## 9.5 Nachträglicher Ausbau der Datenbank

Ein wesentliches Element des RDF-Datenmodells ist, das – abseits von Regelungen in den verwendeten Ontologien – grundsätzlich jede Aussage (also beliebige Tripel) erlaubt ist.<sup>95</sup> Selbst widersprüchliche Aussagen könnten problemlos in einer RDF-basierten Datenbank koexistieren. Es ist folgerichtig problemlos möglich, beliebige neue Aussagen in einen bestehenden Datenbestand aufzunehmen. Probleme entstehen daraus voraussichtlich nur, wenn die Abfrage auf den Datenbestand so formuliert sind, dass die neuen Informationen zu fehlerhaften Ergebnissen führen können.

Die Wahrscheinlichkeit solcher Probleme ist abhängig von der Komplexität der Abfragen: Inferenzen (Schlussfolgerungen) aus der Verkettung von Aussagen und/oder den in den verwendeten Ontologien maschinenlesbar formulierten Rahmenbedingungen können durch neue Tripel vergleichsweise subtil verändert werden. Die coref-Ontologie enthält ihrerseits keine impliziten Aussagen über die mit ihrer Hilfe verknüpften Ressourcen (repräsentiert durch URI). Bei der Auswahl der Prädikate für die Zuordnung der Anreicherungsressourcen zu den bibliographischen Ressourcen sollte dieser Aspekt insofern beachtet werden, dass die in den Ontologien der Prädikate formulierten Aussagen über die Subjekte und Objekte, die durch die Prädikate verknüpft werden, auf Kompatibilität mit den im Kontext des Anreicherungsdienstes herrschenden Rahmenbedingungen geprüft werden müssen.

### 9.5.1 Integration von FRBR-Aussagen

Aus den vorhergehenden Ausführungen insbesondere im Kapitel 8 sollte erkennbar sein, dass eine über die recht simple Koordination von Identifikatoren per coref-Ontologie hinausgehende Integration des FRBR-Modells und der darin vorgesehenen Relationen eine Steigerung der Trefferquote und Effektivität des Anreicherungsdienstes bewirken würde. Als Basis für die Integration entsprechender Tripel steht das Vokabular «RDA Relationships for Works, Expressions, Manifestations, Items» bereit.<sup>96</sup> Dieses definiert Properties bzw. Prädikate, welche die FRBR-Beziehungen repräsentieren.

Tatsächlich stehen von jedem Prädikat mehrere Versionen bereit, die sich untereinander dadurch unterscheiden, welche Anforderungen («rdfs:domain») sie an die Subjekte haben mit denen sie verknüpft werden. Für jedes Relationen-Prädikat gibt es eine Version, die in dieser Hinsicht keinerlei Aussagen trifft sowie weitere, die als Subjekte entweder FRBR-Werke, FRBR-Entitäten oder FRBR-Manifestationen voraussetzen. Welche Variante in einer Ausbaustufe des Anreicherungsdienstes zum Einsatz

---

<sup>95</sup> Vgl. den in Abschnitt 9.1 zitieren AAA-Slogan (siehe dazu Allemang und Hendler, 2008, S. 7 f.).

<sup>96</sup> Vgl. <http://metadataregistry.org/schema/show/id/13.html>.

---

kommt, hängt von der Einschätzung der Zuverlässigkeit der Identifikation der Entitäten ab: Je vorsichtiger diese ausfällt, desto eher sollte die Variante ohne implizite Aussagen über Subjekte gewählt werden.

### 9.5.2 Metadaten über die Anreicherungsressourcen

Zusätzliche Informationen über die Anreicherungsressourcen selbst können zu einer verbesserten Nutzbarkeit führen (vgl. Abschnitt 9.4.3). Sie dürften i. d. R. entweder aus einem genuinen Ausbau der verfügbaren Rohdatenquelle oder aus einer gründlicheren Analyse und erweiterten Verarbeitung dieser Rohdaten erwachsen. In Anbetracht der insgesamt relativ geringen Bedeutung der Anreicherungsressourcen aus sich selbst heraus hat es erscheint es allerdings unwahrscheinlich, dass jemand anderes als der ursprüngliche Lieferant Metadaten über sie beisteuert. Es erscheint daher nicht notwendig, diese Variante der Ergänzung von Daten in Betracht zu ziehen.

## 9.6 Update-Verfahren

Aktualisierungen des Datenbestandes können unterschiedlich kompliziert sein in Abhängigkeit davon, wie die einzelnen Datenbestände selbst aktualisiert oder zugänglich gemacht werden.

Die optimale Lösung sind inkrementelle Updates, die lediglich die seit der letzten Lieferung geänderten Daten umfassen. Neu hinzugekommene Tripel<sup>97</sup> können einfach im Triple Store ergänzt werden, Änderungen oder Löschungen können innerhalb des gleichen Teilbestandes (bzw. Named Graph; vgl. Abschnitt 9.4.4) ebenfalls direkt umgesetzt werden.

Ist in einem anderen Teilbestand bereits ein Tripel mit identischer Aussage enthalten, so ist dies unschädlich. Sind mit beiden Teilbeständen unterschiedliche Nutzungsbedingungen verknüpft, kann es Sinn machen, an dieser Stelle das mit stärkeren Beschränkungen versehene Tripel zu entfernen bzw. gar nicht erst hinzuzufügen. Festzustellen ist allerdings, dass die Bandbreite der möglichen Bedingungen diesen Vergleich zumindest in der Theorie schwierig erscheinen lässt. Ob die Praxis dieses Problem bestätigt, oder ob in der Realität nur eine kleine Auswahl von Bedingungen Bedeutung erlangt, bleibt abzuwarten. Das Problem an sich ist allerdings nicht auf den Moment des Imports beschränkt: Für die konkrete Verwendung (Anzeige) der Anreicherungsressourcen wird eine automatische Auswertung in jedem Fall notwendig sein, wenngleich ein qualitativer Vergleich dabei nicht zwingend erforderlich ist.

Bei Änderungen oder Löschungen von Tripeln, zu denen Duplikate existieren, hängen die Konsequenzen für die Duplikate davon ab, inwieweit der Grund für die Änderung / Löschung erkennbar ist und wie groß das Ver-

---

<sup>97</sup> Wenn im Verlauf dieses Abschnitts von Tripeln gesprochen wird, so sind damit die Tripel gemeint, die aus einer Analyse und Zerlegung der Rohdaten ermittelt wurden – unabhängig davon, in welchem Datenformat diese Rohdaten tatsächlich geliefert wurden.

---

trauen in den jeweiligen Lieferanten ist. Wird beispielsweise ein Tripel geändert, weil die Anreicherungsressource im Objekt des Tripels eine neue URL (und damit URI) erhalten hat, sollten andere Tripel eines anderen Lieferanten, die noch auf die veraltete URL verweisen, mit geändert bzw. entfernt werden, damit dem Endbenutzer durch den Anreicherungsdienst keine «toten» Links angeboten werden. Wie wahrscheinlich es ist, dass derartige Änderungsinformationen überhaupt ausgeliefert werden und noch dazu mit maschinenlesbarer Erläuterung muss die Praxis zeigen.

Periodische Vollarzüge sind etwas aufwendiger in der Prozessierung, lassen sich gegenüber dem vorhandenen Stand aber in inkrementelle Updates umrechnen. Alternativ könnte der alte Teilgraph entfernt und durch einen vollständigen Neu-Import ersetzt werden.

Bei diesem Vorgehen würden allerdings keine Informationen über Duplikate in anderen Teilgraphen erkennbar werden. Ob dies von Bedeutung ist, kann erst in der Praxis beantwortet werden.

Problematisch sind Datenlieferanten, die ihre Daten zwar per API für eine Abfrage und lokale Speicherung bereitstellen, aber keinen Gesamtexport und keine Möglichkeit zur gezielten Abfrage von Änderungen bieten.<sup>98</sup> Sobald ein neuer Identifikator in die Datenbank aufgenommen wird, kann ein solcher Dienst befragt werden. Änderungen in dessen Daten werden jedoch nur sichtbar, wenn periodisch die API für alle in Frage kommende Identifikatoren aufgerufen wird. Ob dies praktikabel ist, hängt von der Performance der API ab. Ein solcher Lieferant wird besser als dynamische Datenquelle in den Anreicherungsdienst einbezogen, indem die ermittelten Daten zu einem Identifikator für eine bestimmte Zeit zwischengespeichert und erst nach Ablauf der Frist wieder aus der eigentlichen Quelle abgerufen werden.

Unter der Annahme, dass bei Abfragen des Dienstes (vgl. Kapitel 10 bzw. insbesondere Abschnitt 10.2.4) mit mehreren Identifikationsmerkmalen alle davon tatsächlich auf die gleiche oder eine hinreichend eng verwandte bibliographische Ressource verweisen, besteht die Möglichkeit, die Abfragen an den Dienst gleichzeitig zum Ausbau der Datenbasis zu verwenden. Dazu würden die eingehenden Anfragen analysiert und die implizit enthaltenen Informationen über Beziehungen zwischen Identifikatoren in zusätzliche coref-Tripel umgesetzt.

---

<sup>98</sup> Es ist unklar, ob dieser Fall überhaupt existiert. Er sei jedoch der Vollständigkeit halber an dieser Stelle beschrieben.

---

## 10 INTEGRATION IN NUTZER-SYSTEME

Die Integration der vom Anreicherungsdienst bereitgestellten Daten in nutzende Systeme kann auf verschiedene Weise erfolgen. Entweder die Anreicherungsressourcen (bzw. die Informationen über die Anreicherungsressourcen) werden periodisch zum Download bereitgestellt, in die lokalen Systeme geladen und von dort ausgelesen. Oder die Anreicherungsressourcen werden im Moment der Anzeige der lokal vorgehaltenen bibliographischen Daten aus dem zentralen Dienst abgerufen und dynamisch in die Anzeige eingefügt.

### 10.1 Bereitstellung per Export und lokaler Import

Eine Methode ist die Bereitstellung der in der Anreicherungsdatenbank (dem Triple Store) gesammelten Daten in einem Gesamt-Export («Dump»). In diesem Fall kommt es im Wesentlichen auf die begleitende Bereitstellung einer Dokumentation der internen Strukturen und der verwendeten Vokabularen an, um die Hürden bei der Auswertung und Übernahme der Daten in Nutzer-Systeme zu minimieren.

In Abhängigkeit von den jeweils zutreffenden Benutzungsbedingungen für die einzelnen Teildatenbestände kann es sein, dass einige davon (identifiziert als «Named Graphs») aus dem Export ausgelassen werden müssen: beispielsweise wenn diese explizit für die Benutzung im Anreicherungsdienst durch dessen Betreiber lizenziert wurden und nicht als Einheit weitergegeben werden dürfen.

Zusätzlicher Aufwand in der laufenden Datenverwaltung entsteht, wenn die Bereitstellung nicht nur über periodische Gesamt-Exporte erfolgen soll, sondern parallel dazu inkrementelle Aktualisierungen erwünscht sind. In diesem Fall müssten alle neu in die Anreicherungsdatenbank einfließenden Tripel registriert werden oder es müsste über eine Software die Differenz zwischen dem aktuellen und dem letzten Stand ermittelt werden. Die Speicherung eines Zeitstempels je Tripel erforderte immensen Aufwand: Identifizierung jedes Tripels über die sogenannte Reification (vgl. Brickley und Guha, 2004, Abschnitt 5.3) oder Behandlung jedes Tripels als eigenen benannten Graphen und Speicherung eines weiteren Tripels (mit dem Graphen bzw. der Reification als Subjekt) für das Datum brächten eine Erhöhung der Tripel-Anzahl um das zwei- bis vierfache mit sich! Sinnvoller erscheint stattdessen ein Mitschreiben der neuen Tripel eines Zeitraums in einer gesonderten Logdatei oder Datenbank, von wo sie dann direkt für den Aktualisierungs-Export entnommen werden können.

Wie genau dann die Verwendung durch die Nutzer-Systeme erfolgt, kann nicht weiter kontrolliert werden.

---

## 10.2 Dynamische Integration

Die dynamische Integration von zusätzlichen Daten im Kontext der Anzeige eines Datengrundbestandes (z. B. eines Katalogdatensatzes) ist inzwischen ein nahezu alltägliches Phänomen im Bibliothekswesen (vgl. Back und Bailey, 2010, S. 76). Die Ergebnisse dieser Prozesse werden als «Mashups» bezeichnet.

Back und Bailey (2010, S. 76 f.) beschreiben zwei grundlegende Konzepte für die technische Realisierung der Zusammenführung der verteilt vorgehaltenen Daten: einerseits eine serverseitige Zusammenführung und andererseits eine Zusammenführung im Client (Browser). Für beide Alternativen führen sie eine Reihe von Vor- und Nachteilen auf, so dass die Auswahl der passenden Methode letztlich vom konkreten Einsatz-Szenario und der Einschätzung der verantwortlichen Personen abhängt. Die vorliegende Literatur zum Thema (u.a. Back und Bailey (2010), Engard (2009), Greenall (2011), Yee (2008)) sowie die berufliche Erfahrung des Autors legen allerdings den Schluss nahe, dass clientseitige Implementierungen vorherrschen (vgl. auch Abschnitt 2.4.5). Primärer Vorteil ist hier, dass die Integration der Anreicherungsdaten die Anzeige der Daten aus der Basis-Applikation (die letztlich immer noch das Hauptinteresse der Nutzer sein sollten) nicht verzögert.

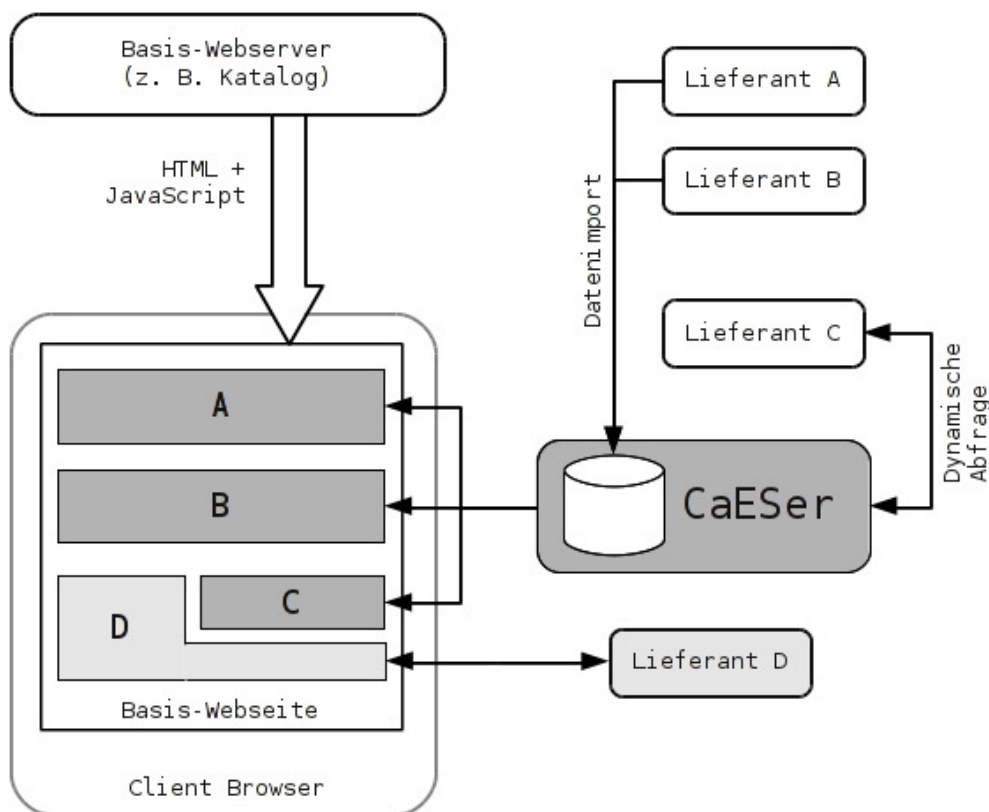
Aufgabe des Anreicherungsdienstes ist es, die notwendigen Schnittstellen bereitzustellen, um eine möglichst breite Nutzung zu erlauben und diese nicht auf Einrichtungen mit besonderer Software-Ausstattung oder besonderen Personalressourcen zu beschränken. Dafür lassen sich drei Methoden identifizieren: erstens die Bereitstellung von vorgefertigten aber konfigurierbaren «Widgets», zweitens die Bereitstellung einer Programmierschnittstelle (API) und drittens die Bereitstellung einer SPARQL-Abfragemöglichkeit<sup>99</sup>.

### 10.2.1 Workflow-Überblick

Das Schaubild in Abbildung 6 zeigt (in Anlehnung an zwei Schaubilder bei Back und Bailey, 2010, S. 77) die wesentlichen Abläufe bei der dynamischen, clientseitigen Integration des Anreicherungsdienstes in eine Applikation. Dies umfasst sowohl die Integration über ein Widget als auch per API. Die Abbildung beschreibt allerdings nicht adäquat die Integration per SPARQL-Abfrage, da der Anreicherungsdienst («CaESer») in der Abbildung tatsächlich als «Dienst» gekennzeichnet ist, der weitere Lieferanten dynamisch integrieren kann und nicht als reine «Datenbank».

---

<sup>99</sup> Auch «SPARQL-Endpoint» genannt.



**Abbildung 6 CaESer - Datenfluss**

Die Daten der Lieferanten A und B werden zu beliebigen Zeitpunkten in die eigentliche Anreicherungsdatenbank importiert und daraus ausgelesen. Lieferant C dagegen wird im Moment einer Anfrage an den Anreicherungs-dienst von diesem dynamisch abgefragt. Die aus den internen und externen Abfragen resultierenden Daten werden in ein einheitliches Format überführt, gemeinsam an die aufrufende Stelle (Browser) übermittelt und dort in die Basis-Webseite einer anderen Applikation integriert.

Die eigentliche Anreicherungsdatenbank kann dabei verwendet werden, um aus den dort gespeicherten Koreferenzen zunächst Alternativen zu dem oder den Identifikatoren in der Anfrage zu ermitteln. Diese alternativen Identifikatoren erlauben es unter Umständen, dynamisch integrierte weitere Lieferanten abzufragen, für die die ursprünglichen Identifikatoren ungeeignet waren oder diese Lieferanten mit mehreren Identifikatoren abzufragen.

Die Basis-Webseite kann zusätzlich beliebige weitere Lieferanten (im Schaubild stellvertretend repräsentiert durch «Lieferant D») dynamisch und unabhängig vom Anreicherungsdienst integrieren. Dies ist aus verschiedenen Gründen sinnvoll. Zum einen ist davon auszugehen, dass einige Lieferanten auf einer direkten, nicht durch einen Proxy vermittelten Integration bestehen werden. Zum anderen wird der Anreicherungsdienst niemals alle möglichen Lieferanten abdecken können – langfristig nicht und erst recht

---

nicht zum Zeitpunkt der Inbetriebnahme. Dies gilt insbesondere für lokale Anreicherungsdatenbanken einzelner Institutionen. Schließlich können technische Gründe (lange Antwortzeiten, geringe Zuverlässigkeit) gegen eine Aufnahme in den Anreicherungsdienst sprechen, deren Relevanz durch den Betreiber der Basis-Applikation anders eingeschätzt wird.

### 10.2.2 Widget

Unter einem «Widget» wird hier folgendes verstanden: Die Kombination eines HTML-«Schnipsels» und des Verweises auf eine oder mehrere JavaScript-Dateien sowie ggf. einer kleinen Menge an JavaScript-Code zur Angabe von Konfigurationsparametern, die gemeinsam in die Vorlage einer dynamisch zu generierenden Webseite eingebaut werden, wobei der JavaScript-Code zum Zeitpunkt der Anzeige der generierten Seite in einem Browser einen Server kontaktiert, Daten abholt und darauf aufbauend die angezeigte Seite an der spezifizierten Stelle modifiziert bzw. ergänzt. Die Aufgabe des Widgets umfasst also explizit nicht nur die Ermittlung der Rohdaten für das Mashup, sondern zusätzlich auch die Aufbereitung dieser Daten für die Anzeige. Widgets sind für Back und Bailey (2010, S. 78) das Mittel der Wahl für die Implementierung von Mashups:

The amount of JavaScript code a librarian needs to write [...] determines both the likelihood of adoption and the maintainability of a given mash-up creation. The less JavaScript code there is to write, the larger the group of librarians who feel comfortable trying and adopting a given implementation. The approach of using HTML widgets hides the use of JavaScript almost entirely from the mash-up creator.

Greenall (2011, Abschnitt Functional Requirements”) ergänzt zusätzlich die Anforderung, dass der JavaScript- und HTML-Code des Widgets möglichst in jeder aktuell genutzten Browser-Variante funktionieren muss.

Selbstverständlich erfordern auch Widgets eine Schnittstelle auf Seiten des Servers (also des Anreicherungsdienstes), der die Daten für das Mashup bereitstellt. Da der JavaScript-Code im Client, der die Daten abfragt und für die Ausgabe aufbereitet, jedoch vom Betreiber des Anreicherungsdienstes entwickelt und über dessen Server bereitgestellt wird, kann die Integration zwischen beiden Software-Komponenten besonders eng ausfallen und können Entwicklungen und Veränderungen jederzeit nach Bedarf vorangetrieben werden.<sup>100</sup> Auch ist die Erstellung einer umfangreichen Dokumentation für eine nur für interne Zwecke gedachte API nicht notwendig.

Die Anforderungen an die Gestaltung der Anzeige der Anreicherungsressourcen andererseits sind – zumindest theoretisch – hoch. Allgemeine Verweise auf Datenlieferanten oder Links zu bestimmten Lieferantenwebseiten

---

<sup>100</sup> Die Ausgabe an sich sollte allerdings mit Blick auf die Endnutzer nicht allzu kurzen Entwicklungszyklen unterworfen werden, um nicht unnütz Verwirrung zu stiften.

---

in Abhängigkeit von der jeweiligen Anreicherungsressource sind ebenso zu beachten wie die Frage, wie beispielsweise mehrere Umschlagabbildungen auf einmal mit möglichst geringem Platzbedarf angezeigt werden können; direkt anzeigbare Daten wie kurze Abstracts sind anders zu behandeln wie Links zu Inhaltsverzeichnissen in PDF-Dokumenten.

### 10.2.3 Programmierschnittstelle (API)

Eine offiziell freigegebene API zur Benutzung durch Dritte erlaubt es den Anwendern, die Abfrage und Ausgabe der Daten flexibel an die eigenen Bedürfnisse anzupassen. Diese Art der Nutzung erfordern jedoch auf Seiten der nutzenden Institution ein erheblich höheres Maß an Vorwissen, Programmierkenntnissen und einmaligen sowie laufenden Wartungsaufwand.

Der Begriff der Schnittstelle umfasst hier mehrere Aspekte: die URL, unter der die Funktionalität angesprochen werden kann; die Benennung möglicher Aufrufparameter und die dafür jeweils akzeptierten Werte; die Struktur und möglichen Inhalte der Antwort. Einmal definiert, sollten sich diese Informationen möglichst selten und erst recht nicht in inkompatibler Art ändern. Sofern größere Änderungen eingebracht werden müssen, sollte die bisherige Version möglichst für einen längeren Zeitraum unverändert verfügbar bleiben. Die neue Variante könnte dazu entweder über eine neue URL oder über einen Versions-Parameter in der Anfrage angesprochen werden.

Der Datenaustausch erfordert eine strukturierte Ausgabe der ermittelten Informationen zu verfügbaren Anreicherungsressourcen, damit die aufrufende Stelle diese Daten für eine Anzeige aufbereiten kann. Grundsätzlich ist die Auswahl eines beliebigen Austauschformats denkbar. Allerdings ist der erwartete typische Anwendungsfall der Aufruf per JavaScript aus dem Browser heraus. Es macht daher Sinn, ein Format zu wählen, das insbesondere für diesen Anwendungsfall gut geeignet ist. Dies findet sich leicht in der sogenannten «JavaScript Object Notation», kurz «JSON» (vgl. JSON). Dieses Format erfreut sich breiter und stetig zunehmender Beliebtheit im Kontext von Web-basierten Anwendungen (vgl. DuVander, 2011). Back und Bailey (2010, S. 78) argumentieren ebenfalls deutlich pro JSON für vergleichbare Einsatzzwecke.<sup>101</sup>

Zudem existieren Spezifikationen für die Serialisierung von RDF-Daten in JSON (Heath und Bizer, 2011, Abschnitt 2.4.2), wodurch die Nutzung des gesammelten Daten vereinfacht wird. Eine Alternative wäre die Verwendung eines existierenden JSON-basierten Schemas für den Transport

---

<sup>101</sup> Back und Bailey (2010, S. 78) schreiben JSON dabei auch die Eigenschaft zu, ein bestimmtes Sicherheitsfeature (vgl. Same origin policy) in Browsern umgehen zu können, was tatsächlich zwangsläufig notwendig ist, um Daten aus dem zentral betriebenen Anreicherungsdienst in verteilt betriebene Applikation integrieren zu können. Leider verwechseln sie dabei das Format JSON mit dem Anwendungsmuster JSONP (vgl. JSONP). Dies ändert jedoch nichts an der grundlegenden Eignung von JSON/JSONP als Mittel zur Lösung der vorliegenden Aufgabe.



---

der auszugebenden Daten. Damit ließe sich der Anreicherungsdienst zusätzlich leicht in vorhandene Anwendungen integrieren. Theoretisch in Frage käme der «SeeAlso»-Standard:

The creation of the SeeAlso linkserver protocol was occasioned by the need to enrich title views in library catalogues of the German Common Library Network (GBV) with links to additional information. However, instead of integrating those links into title records and tailoring the presentation to our specific OPAC software, we decided to create a general linkserver Web service. (Voß, 2008b)

Die SeeAlso-Spezifikation (Voß, 2009) enthält keine inhärenten Beschränkungen hinsichtlich der abzurufenden Informationen, wenngleich der Text insgesamt nahe legt, dass der direkte Transport von Texten in einer Länge jenseits weniger Worte (z. B. Abstracts oder Rezensionen) nicht vorgesehen war. Die Spezifikation beschreibt zudem – passend zur internen Datenhaltung des projektierten Anreicherungsdienstes – einen Mechanismus zum Transfer von RDF-Tripeln, wobei jedoch gerade die Möglichkeit zum Transfer von qualifizierenden Attributen zu Literalen (als Objekte der Tripel) entfällt.<sup>102</sup>

SeeAlso sieht vor, dass Anfragen jeweils genau einen Identifikator enthalten. Diese Beschränkung müsste für eine Nutzung im vorliegenden Kontext aufgehoben werden. Problematischer ist, dass SeeAlso davon ausgeht, dass eine Anfrage jeweils nur einen Typus von Anreicherungsressourcen zurückliefert und keine Strukturen enthält, um diesen Typus zu beschreiben. Da Tags oder Umschlagabbildungen jeweils völlig verschiedene Arten der Anzeige erfordern, ist die Unterscheidung jedoch notwendig. Auch sieht die Spezifikation keine Möglichkeit vor, um in einer Antwort Anreicherungen verschiedenen Identifikatoren zuzuordnen. Dies wird umso kritischer, je stärker eine FRBRisierung zum Einsatz kommt.

Im Endeffekt ergeben sich aus dieser Eignungsprüfung zwei Erkenntnisse. Erstens wird ein eigenes Antwort-Format des Anreicherungsdienstes komplexer ausfallen als das SeeAlso-Format, damit die Zuordnung von Anreicherungsressourcen zu Identifikatoren, die Kennzeichnung unterschiedlicher Typen von Anreicherungsressourcen sowie die Angabe von Metadaten zu einzelnen Anreicherungsressourcen möglich ist. Zweitens könnte der Anreicherungsdienst mit relativ geringem Aufwand zusätzlich SeeAlso unterstützen. Dafür wäre im Wesentlichen die Einrichtung von individuellen Basis-URLs für die verschiedenen Typen von Anreicherungsressourcen notwendig. Im Gegenzug würde die Anzahl der notwendigen Zugriffe auf den Anreicherungsdienst deutlich ansteigen: Anzahl Identifikatoren × Anzahl Typen von Anreicherungsressourcen.

---

<sup>102</sup> «Please note that literal values with language tag or data types (typed literals) are not supported.» (Voß, 2009).

---

#### 10.2.4 Konfigurations- und Aufrufparameter für Widgets und API

Wenngleich die selbst erstellten Widgets nicht zwangsläufig die offiziell zur Verfügung gestellten Schnittstellen nutzen müssen, macht dies zumindest mittel- bis langfristig Sinn, da es doppelten Pflegeaufwand für zwei Schnittstellen vermeidet. Daraus folgt, dass die Konfigurationsoptionen des Widgets (soweit sie nicht die Darstellung der Daten durch Bearbeitung innerhalb des Widgets beeinflussen) eine Entsprechung in den Aufrufparametern der Schnittstelle haben müssen oder zumindest sinnvolle Kandidaten dafür sind.

Viele bibliographische Ressourcen enthalten mehr als ein Identifikationsmerkmal: neben einer ISBN dürften sich z. B. Identifikatoren der Nationalbibliothek bzw. Nationalbibliographie oder eines Katalogisierungsverbundes finden oder (bei E-Books) DOI. Zumindest in der Theorie erhöht die Anzahl der Identifikationsmerkmale in der Anfrage die Chancen auf einen Treffer, so dass ein solches Vorgehen zulässig sein sollte. Es ist anzunehmen, dass sich dieser Ansatz umso mehr auszahlen wird, je weiter eine «FRBRisierung» der Datenbestände im Client-System und/oder im Anreicherungsdienst voranschreitet, da sich damit die Zahl der verknüpften «öffentlichen» Identifikationsmerkmale wie ISBN oder DOI (in Abgrenzung zu primär im unmittelbaren bibliothekarischen Umfeld bekannten Identifikatoren wie Verbundkatalog-IDs) deutlich erhöhen wird. Wichtig ist dabei, das grundsätzlich auch der Typ des Identifikators mit angegeben wird, da sich diese Information mangels eindeutiger Bestandteile nur mit einer erheblichen Fehlerquote durch die Analyse der Zeichenketten ermitteln lässt.

Nach wie vor gibt es in Bibliotheken Vorbehalte gegen die Anreicherung der Trefferanzeige von Katalogen mit bestimmten Typen von Anreicherungsressourcen. Dies betrifft insbesondere Informationen mit wertendem oder werbendem Charakter. Zu ersteren gehören Rezensionen (vgl. dazu auch Abschnitt 2.3). Letzteres meint beispielsweise die Anzeige von Klappentexten zu Büchern.

Weitere Vorbehalte gibt es grundsätzlich gegenüber der Anzeige von Links zu externen Webseiten und insbesondere zu solchen, die zu kommerziellen Anbietern führen.

Unabhängig davon, ob diese Vorbehalte gerechtfertigt sind, kann ihre Existenz nicht ignoriert werden, da sie zu einer vollständigen Ablehnung des Dienstes führen können. Daher sind Einstellungsmöglichkeiten vorzusehen, die den gezielten Ausschluss 1) bestimmter Typen von Anreicherungsressourcen, 2) kommerzieller Anbieter allgemein sowie 3) einzeln zu benennender Anbieter erlauben (Negativlisten). Fall 1 kann über die verwendeten Prädikate in den Tripeln umgesetzt werden. Die Fälle 2 und 3 lassen sich anhand von Informationen über die Named Graphs in der Anreicherungsdatenbank abbilden. Ein grundsätzlicher Ausschluss von externen Links andererseits erscheint überzogen.

---

Wenn bestimmte Informationen seitens des Anreicherungsdienstes nicht ermittelt oder übertragen werden müssen, sollte dies zumindest geringfügig positive Auswirkungen auf die Performance des Systems haben. Dies gilt insbesondere für den Ausschluss von dynamisch integrierten Anbietern, da deren Abfrage mehr Zeit in Anspruch nimmt, als die Abfrage des Triple Stores.

Grundsätzlich plausibel erscheint eine Filterung der Anreicherungen hinsichtlich der verwendeten Sprache. Ohne weitere Angabe sollten alle Sprachen berücksichtigt werden; Anwendern könnte jedoch die Möglichkeit eingeräumt werden, die Ergebnisse auf eine oder mehrere Sprachen einzuschränken (Positivliste). Ob eine solche Konfigurationseinstellung bzw. ein entsprechender Parameter sinnvoll ist, wird sich jedoch erst während des Aufbaus des Dienstes zeigen, wenn erkennbar wird, für welchen Prozentsatz an erfassten Anreicherungsressourcen die notwendige Information vorliegt.

Im Falle einer sich weiterentwickelnden FRBRisierung der Daten im Anreicherungsdienst erscheint die Beschränkung der Ausgabe auf Anreicherungsressourcen innerhalb einer einzigen Manifestation oder Edition denkbar. Theoretisch könnte hierbei sogar noch nach Typen von Anreicherungsressourcen unterschieden werden: beispielsweise die Abfrage von Schlagwörtern auf der Werk-Ebene, von Inhaltsverzeichnissen jedoch auf der Manifestations-Ebene. Welcher Ansatz sinnvoll und vor allem praktikabel ist, kann jedoch nur anhand konkreter Erfahrungen mit realen Daten beurteilt werden.

Schließlich erscheint – aufgrund der Ausführungen zu SeeAlso im Abschnitt 10.2.3 – noch die Einrichtung von Varianten der Schnittstelle sinnvoll, die auf die Ausgabe jeweils eines bestimmten Typs an Anreicherungsressourcen beschränkt sind. Statt über einen Parameter müsste diese Beschränkung allerdings über eine modifizierte URL ausgedrückt werden, um die Kompatibilität zur SeeAlso-Spezifikation zu erhalten.

Alle vorstehend beschriebenen Aufrufparameter mit Ausnahme mindestens eines Identifikators sowie der Angabe des JSONP «Paddings» (vgl. JSONP ) sind optional, um eine Benutzung so unaufwendig wie möglich zu halten.

### 10.2.5 SPARQL

Eine SPARQL-Abfragemöglichkeit ist Bestandteil aller Triple Stores und kann somit durch den Betreiber der Anreicherungsdienstes fast ohne Aufwand bereitgestellt werden. Im Hinblick auf den Anspruch des «Linked Open Data»-Ansatzes, Daten möglichst durchgängig mit Mitteln bereitzustellen, die auf offenen Standards beruhen (vgl. A. Pohl, 2010, S. 6 f.) ist dies zudem zweifellos zu begrüßen. Allerdings fehlt bei diesem Vorgehen jegliche «Isolationsschicht» zwischen den Anwendern und den Daten. Änderungen in den Datenstrukturen können daher zuvor funktionierende Abfragen plötzlich fehlschlagen lassen. Sind in der Anreicherungsdatenbank Teilbestände enthalten, die nur unter bestimmten Bedingungen vom Anrei-

---

cherungsdienst verwendet werden dürfen (z. B. nur punktueller Zugriff aber kein Massen-Download), so entfallen beim SPARQL-Zugriff sämtliche Sicherungsmethoden, die in eine API eingebaut werden können (z. B. eine Beschränkung auf eine bestimmte maximale Anzahl von Zugriffen pro Zeiteinheit und nutzender Institution).

Ein SPARQL-Endpoint sollte daher nur angeboten werden, wenn die Datenqualität einigermaßen gesichert ist (Verwendung langfristig gültiger URIs und abgeschlossene Evaluierung des verwendeten RDF-Vokabulars). Alternativen bestehen im Angebot von Datenexporten (unter Auslassung kritischer Daten; vgl. Abschnitt 10.1) oder in der Bereitstellung eines SPARQL-Endpoints über einen separaten Triple Store mit den für den Export ausgewählten Teilbeständen.

Da die SPARQL-Abfrage direkt auf dem Triple Store beruht, können die Ergebnisse dieser Abfragen – anders als die Abfragen per API – keine Daten von dynamisch durch den Anreicherungsdienst selbst integrierten Lieferanten enthalten.

---

## 11 IMPLEMENTIERUNG

Die Diskussionen der einzelnen Aspekte hat gezeigt, dass für eine (der Theorie nach) optimale Lösung erheblicher Aufwand zu investieren wäre. Zudem wurden einige praktische Aspekte wie die Auswahl der passenden Softwareplattform oder der Aufwand für die Auswertung der Rohdaten noch gar nicht näher betrachtet. Eine Implementierung sollte daher in Form eines inkrementellen Auf- und Ausbaus der angestrebten Funktionalität umgesetzt werden.

Die Verwendung eines Triple Store bringt große Flexibilität beim Aufbau der Datenbank und vermeidet den Aufwand zum Entwurf eines normalisierten Datenbank-Schemas. Im Gegenzug sind allerdings die RDF-Strukturen zu prüfen und auf effiziente Abfrage-Möglichkeiten per SPARQL-Abfragen zu untersuchen. Um Erfahrungen mit dieser Technik und der effizienten Abfrage der Daten zu sammeln, sowie zugleich ein grundlegendes Software-Framework für die Auswertung von Datenbeständen und den Import zu entwickeln, sollten in einem ersten Schritt einige möglichst heterogene Datenquellen ausgewählt und testweise importiert werden.

Sinnvollerweise sollte dazu auf Daten zurückgegriffen werden, die aufgrund einer entsprechenden Verzichtserklärung des Rechteinhabers uneingeschränkt zur Verfügung stehen.

Im Produktionsbetrieb werden drei wesentliche Arbeitsgebiete zu unterscheiden sein. Einmal die Pflege und Weiterentwicklung des Basis-Systems und der Schnittstellen. Des weiteren die Suche nach und Einwerbung von neuen Datenbeständen sowie die Prüfung von Lizenzen solcher Datenbestände. Und schließlich die Entwicklung und Pflege von Programmen, die die eingehenden Datenlieferungen analysieren und für die Aufnahme in die Datenbank aufbereiten.

Bezüglich der Festlegung und Dokumentation einer API ist festzustellen, dass zum Ersten ein politisches Motiv für den zuvor beschriebenen Anreicherungsdienst in der Bereitstellung eines Angebots gerade auch für kleinere Bibliotheken ohne EDV-Ressourcen besteht und zum Zweiten eine solche API auch eine Testumgebung benötigt. Infolgedessen sollte zunächst die Erstellung eines Widgets im Vordergrund der Aktivitäten stehen. Die dafür zu erstellende Schnittstelle kann parallel auch anderen Anwendern zur Verfügung gestellt werden, jedoch mit der Einschränkung, dass die Stabilität bis zu einer offiziellen Freigabe nicht gewährleistet wird.

Ob bzw. zu welchem Zeitpunkt ein SPARQL-Endpoint oder RDF-Exporte bereitgestellt werden, ist sowohl eine technische wie auch eine politisch-strategische Frage. Liegt der Fokus eher auf der Positionierung des Anreicherungsdienstes als Produkt, so ist die Bereitstellung der aufbereiteten Daten nicht erforderlich oder sogar – als Material für Konkurrenten – schädlich. Durch einen solchen Schritt gingen allerdings auch der Zugriff auf Daten unter entsprechenden Lizenzen und die Bindung an die «Linked Open Data» Entwicklungsgemeinschaft verloren. Nach einer internen Testpha-

---

se mit uneingeschränkt nutzbaren Daten sollte daher die Öffnung und Preisgabe der Daten erfolgen.

---

## 12 FAZIT

Ein Projekt zur Implementierung eines Dienstes entlang der dargestellten Überlegungen erscheint insgesamt umsetzbar. Nicht alles, was aus funktionaler Sicht wünschenswert ist (insbesondere die FRBRisierung der Daten), kann dabei allerdings von Anfang an und/oder aus eigener Kraft realisiert werden.

Zu den umfangreichsten und damit wichtigsten Quellen für Anreicherungsressourcen gehören die Verleger (die den Buchhandel zu Marketingzwecken mit diesen Informationen versorgen) sowie die Kataloge von Verbänden, Nationalbibliotheken und sonstigen großen Bibliotheken, die selbst Anreicherungsressourcen (Scans von Inhaltsverzeichnissen, Schlagwörter usw.) produzieren. Während Bibliotheken wie die Library of Congress bereits seit Jahren umfangreiche Datenlieferungen erhalten (vgl. Byrum, 2005) ist unklar, ob diese Daten auch für die Nutzung in einen Anreicherungsdienst eingeworben werden könnten.

Auch die unentgeltliche Nutzung von durch Bibliotheken erstellten Anreicherungsressourcen ist keineswegs bereits gesichert, zumal sich die öffentliche Open Data Diskussion bislang weitgehend auf offene bibliographische Daten beschränkt (vgl. etwa Bérard (2011), A. Pohl (2010) oder Why Open Up Bibliographic Data; wobei die bibliographischen Daten immerhin auch Schlagwörter und Notationen umfassen). So geben derzeit einzelne Bibliotheken des hbz-Verbundes jeweils ihren eigenen Teilbestand als «Linked Open Data» frei, weil nicht alle Verbundteilnehmer mit der Freigabe des vollständigen Verbundkatalogs einverstanden waren. Es steht zu hoffen, dass sich diese Einstellung im Laufe der Zeit ändert, weil sich erst durch die Kooperation zwischen Bibliotheken aller Art die lokalen Investitionen durch eine breite Nutzung wirklich rentieren. Aus Sicht einer einzelnen Bibliothek mögen allerdings Anreicherungsressourcen wie gescannte Inhaltsverzeichnisse einen größeren Wert darstellen als einfache Katalogisate (die es in vielfachen Dubletten gibt und deren Massenerwerb als Fremddaten oder per Verbundkatalogisierung lange etablierte Praxis ist), so dass die Vorbehalte gegen eine freie Weitergabe hier noch größer sein könnten. Firmen wie LibraryThing oder Bowker können einzelne Bibliotheken keine Alternativen gegenüberstellen, in Kooperation jedoch könnte ein wertvolles Angebot entstehen.

Ein wichtiger Aspekt der Thematik ist dabei auch die Aufklärung der Bibliotheken, welche Daten sie unter welchen Rahmenbedingungen aus ihren Beständen legal extrahieren und verfügbar machen dürfen. Das Hochschulbibliothekszenrum des Landes NRW (hbz) hat dazu einen Leitfaden erstellen lassen (s. Kreutzer, 2011), der – trotz seines in dieser Hinsicht eventuell missverständlichen Titels – auch den Bereich der Kataloganreicherung behandelt und viele Fragen beantwortet.

Aufgrund des mit der Implementierung verbundenen Aufwands stellt sich des weiteren die Frage, ob es alternative Ansätze für die Erreichung des

---

zugrunde liegenden Ziels gibt. Dazu lassen sich zwei Möglichkeiten identifizieren.

Sofern das Culturegraph Projekt sein Ziel erreicht, und in großem Umfang ein Mapping von Identifikatoren von bibliographischen Ressourcen in einer Produktionsumgebung bereitstellt, könnte diese Teilfunktion (die co-ref-Bundles) aus der Datenbank des Anreicherungsdienstes entfernt werden. Übrig blieben hier dann die Zuordnungen von Anreicherungsressourcen zu Identifikatoren. Der Anreicherungsdienst würde in diesem Fall zunächst eine Abfrage an Culturegraph stellen, um die ko-referenten Identifikatoren zu ermitteln und dann im zweiten Schritt die zugehörigen Anreicherungsressourcen aus seiner eigenen Datenbank zu ermitteln.

Die zweite Möglichkeit ist wesentlich radikaler. Wieder würde Culturegraph für das Mapping von Identifikatoren eingesetzt. An die Stelle einer eigenen Datenbank im Anreicherungsdienst träte jedoch eine enge Kooperation mit dem Open Library Projekt, dessen Datenbank um alle auffindbaren Anreicherungsressourcen ergänzt würden. Die sequentielle Abfrage der beiden Dienste könnte entweder durch einen zentral betriebenen «Vermittlungsdienst» erfolgen, der seinerseits per JavaScript aus den Client-Browsern abgefragt würde. Oder aber die gesamte Funktionalität wandert in eine JavaScript-Bibliothek, die aus dem Browser heraus zuerst Culturegraph und dann Open Library (und theoretisch weitere Quellen) abfragt. In letzterem Fall müsste kein zentraler Dienst mehr betrieben werden. Dieses Vorgehen könnte jedoch nur in Kooperation mit (und in Abhängigkeit von) dem Open Library Projekt erfolgen und setzt voraus, dass man dort an der Integration der Anreicherungsressourcen (bzw. von deren Nachweisen) interessiert ist bzw. dass die Lizenzbedingungen dies jeweils erlauben.



---

## 13 LITERATURVERZEICHNIS

„Actionable ISBN Application Introduced in Germany“ (2011). In: Information Standards Quarterly 23.1 (Winter), S. 39. ISSN: 1041-0031. DOI: 10.3789/isqv23n1.2011.10

Allemang, Dean und James A. Hendler (2008). Semantic Web for the Working Ontologist. Modeling in RDF, RDFS and OWL. Morgan Kaufmann Publishers. xvii, 330 S. ISBN: 978-0-12-373556-0.

Arbeitsgruppe Kooperative Verbundanwendungen der Arbeitsgemeinschaft der Verbundsysteme (AG KVA), Hrsg. (1. Feb. 2010). Elektronische Adressen von Kataloganreicherungsobjekten in den zugehörigen Metadaten. 5 S. URL: [http://www.dnb.de/SharedDocs/Downloads/DE/DNB/wir/agkva\\_prax\\_kat\\_anr\\_2011\\_web.pdf](http://www.dnb.de/SharedDocs/Downloads/DE/DNB/wir/agkva_prax_kat_anr_2011_web.pdf) (besucht am 13.04.2012).

Back, Godmar und Annette Bailey (2010). „Web Services and Widgets for Library Information Systems“. In: Information Technology and Libraries 29.2 (June), S. 76–86. ISSN: 0730-9295. URL: <http://www.ala.org/lita/ital/files/29/2/back.pdf> (besucht am 13.04.2012).

Bérard, Raymond (2011). „Free Library Data?“ In: LIBER Quarterly 20.3/4, S. 321—331. ISSN: 1435-5205. URL: [http://liber.library.uu.nl/publish/issues/2010-3\\_4/index.html?000512](http://liber.library.uu.nl/publish/issues/2010-3_4/index.html?000512) (besucht am 04.04.2012).

Bates, Marcia J. (1. Juni 2003). Task force recommendation 2.3 research and design review : improving user access to library catalog and portal information : final report (version 3). Library of Congress. 58 S. URL: <http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf> (besucht am 04.04.2012).

Bericht zur Lage der Bibliotheken 2010 (21. Okt. 2010). Deutscher Bibliotheksverband e.V. (dbv). 12 S. URL: [http://www.bibliothekerverband.de/fileadmin/user\\_upload/DBV/publikationen/Bericht\\_zur\\_Lage\\_der\\_Bibliotheken\\_2010\\_01.pdf](http://www.bibliothekerverband.de/fileadmin/user_upload/DBV/publikationen/Bericht_zur_Lage_der_Bibliotheken_2010_01.pdf) (besucht am 13.04.2012).

Berners-Lee, Tim (27. Juli 2006). Linked Data. Version 2009/06/18 18:24:33. URL: <http://www.w3.org/DesignIssues/LinkedData.html> (besucht am 04.04.2012).

- 
- Berners-Lee, Tim, James Handler und Ora Lassila (2001). „The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities“. In: Scientific American 284.5, S. 34–43. ISSN: 0036-8733. URL: <http://www.scientificamerican.com/article.cfm?id=the-semantic-web> (besucht am 04.04.2012).
- Bizer, Christian (5. März 2010). Overview of Named Graphs. Presentation to the W3C Provenance XG. URL: <http://www4.wiwiiss.fu-berlin.de/bizer/pub/Bizer-NamedGraphs-ProvXG.pdf> (besucht am 13.04.2012).
- Bizer, Christian, Richard Cyganiak und Tobias Gauß (2007). „RDF Book Mashup. From Web APIs to a Web of Data“. In: 3rd Workshop on Scripting for the Semantic Web (SFSW2007) (Innsbruck, Österreich, 6. Juni 2007).
- Brickley, Dan und R. V. Guha, Hrsg. (10. Feb. 2004). RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation 10 February 2004. URL: <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/> (besucht am 04.04.2012).
- Brown, Glenn Otis (25. Mai 2004). Announcing (and explaining) our new 2.0 licenses. Creative Commons. URL: <http://creativecommons.org/weblog/entry/4216> (besucht am 04.04.2012).
- Byrum, John D. (2005). „Recommendations for urgently needed improvement of OPAC and the role of the National Bibliographic Agency in achieving it“. In: World Library and Information Congress: 71th IFLA General Conference and Council (Oslo, Norwegen, 14.–18. Aug. 2005). URL: <http://archive.ifla.org/IV/ifla71/papers/124e-Byrum.pdf> (besucht am 04.04.2012).
- Byrum, John D. und David W. Williamson (2006). „Enriching traditional cataloging for improved access to information. Library of congress tables of contents projects“. In: Information Technology and Libraries 25.1, S. 4–11. ISSN: 0730-9295. URL: <http://www.ala.org/lita/ital/files/25/1/byrum.pdf> (besucht am 13.04.2012).
- Campbell, Lorna M. und Sheila MacNeill (2010). The Semantic Web, Linked and Open Data. A Briefing Paper. 6 S. URL: [http://wiki.cetis.ac.uk/images/1/1a/The\\_Semantic\\_Web.pdf](http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf) (besucht am 04.04.2012).

---

Cataloging Enrichment Initiative, Hrsg. (19. Apr. 2004). Draft File Content Element List. URL: <http://www.loc.gov/standards/catenrich/catenrich-elements.html> (besucht am 20.02.2011).

Coffman, Steve (1999). „The Response to ”Building Earth’s Largest Library”“. In: Searcher 7.7, S. 29–32. ISSN: 1070-4795.

Cool URIs for the Semantic Web (3. Dez. 2008). W3C Interest Group Note 03 December 2008. W3C. Sauermann, Leo und Richard Cyganiak, Hrsg. URL: <http://www.w3.org/TR/2008/NOTE-cooluris-20081203/> (besucht am 04.04.2012)..

Coyle, Karen (12. Aug. 2011). Models of bibliographic data. Blog-Eintrag. URL: <http://kcoyle.blogspot.com/2011/08/models-of-bibliographic-data.html> (besucht am 04.04.2012).

Czerwinski, Silvia (2009). „Funktionen des Bibliothekskatalogs. Eine Diskursanalyse der bibliothekarischen Fachkommunikation vom OPAC bis zum Katalog 2.0“. Masterarbeit. FH Köln, Institut für Informationswissenschaft. 97 S.

Czerwinski, Silvia und Jakob Voß (2010). „LibraryThing – die kollaborative Bibliothek 2.0“. In: Handbuch Bibliothek 2.0. Hrsg. von Julia Bergmann und Patrick Danowski. De Gruyter Saur, S. 333–351. ISBN: 978-3-11-023209-7 978-3-598-023210-3.

Davis, Ian (17. Juli 2009). Linked Data and the Public Domain. URL: <http://blogs.talis.com/nodalities/2009/07/linked-data-public-domain.php> (besucht am 04.04.2012).

Denton, William (13. Juni 2006). Example 1: Harry Potter and the Goblet of Fire. Blog-Eintrag. URL: <http://www.frbr.org/2006/06/13/eg-1> (besucht am 04.04.2012).

Denton, William (2007). „FRBR and the History of Cataloging.“ In: Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools. Hrsg. von Arlene G. Taylor. Westport, Connecticut: Libraries Unlimited, S. 35–57. ISBN: 9781591585091. hdl: <http://hdl.handle.net/10315/1250>. (besucht am 13.04.2012).

---

Designing URI Sets for the UK Public Sector (9. Okt. 2009). A report from the Public Sector Information Domain of the CTO Council's cross-Government Enterprise Architecture. Version 1.0. Chief Technology Officer Council. 13 S. URL: <http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-uri-sets-uk-public-sector.pdf> (besucht am 04.04.2012).

Dickey, Timothy J. (2008). „FRBRization of a Library Catalog. Better Collocation of Records, Leading to Enhanced Search, Retrieval, and Display“. In: Information Technology and Libraries 27.1 (March), S. 23–32. ISSN: 0730-9295. URL: <http://www.ala.org/lita/ital/files/27/1/dickey.pdf> (besucht am 13.04.2012).

Dillon, Martin und Patrick Wenzel (1989). „Enhanced bibliographic record retrieval experiments“. In: OCLC Newsletter 181, S. 13–14. ISSN: 0163-898X.

DuVander, Adam (25. Mai 2011). 1 in 5 APIs Say "Bye XML". URL: <http://blog.programmableweb.com/2011/05/25/1-in-5-apis-say-bye-xml/> (besucht am 04.04.2012).

Eberhardt, Joachim (2006). „Zum moralischen Status des Catalogue Enrichment“. In: LIBREAS. Library Ideas 1. lfd. Nr. 4. ISSN: 1860-7950. URL: [http://www.ib.hu-berlin.de/~libreas/libreas\\_neu/ausgabe4/003ebe.htm](http://www.ib.hu-berlin.de/~libreas/libreas_neu/ausgabe4/003ebe.htm) (besucht am 04.04.2012).

Emanuel, Jenny (2009). „Next Generation Catalogs. What Do They Do and Why Should We Care?“. In: Reference & User Services Quarterly 49.2, S. 117–120. ISSN: 1094-9054. URL: <http://rusa.metapress.com/content/n5547412578742u5/> (besucht am 13.04.2012).

Eversberg, Bernhard (4. März 2002). Zur Theorie der Bibliothekskataloge und Suchmaschinen. Version vom 06.07.2005. URL: <http://www.allegro-c.de/formate/tks.htm> (besucht am 04.04.2012).

Eversberg, Bernhard (9. März 2004a). Sachliche Erschließung. Eine Aufgabe mit vielen Facetten. Version vom 29.04.2005. URL: <http://www.allegro-c.de/formate/se.htm> (besucht am 04.04.2012).

Eversberg, Bernhard (22. Sep. 2004b). Zur Zukunft der Katalogisierung. jenseits RAK und AACR2. Version vom 15.07.2008. URL: <http://www.allegro-c.de/formate/zk.htm> (besucht am 13.04.2012).

- 
- Flimm, Oliver (18. Juni 2008). Zentrale Kataloganreicherung. Blog-Eintrag. URL: <http://blog.openbib.org/2008/06/18/zentrale-kataloganreicherung/> (besucht am 04.04.2012).
- Folksonomy. Bearbeitungsstand: 15. Juli 2011, 16:41 UTC. In: Wikipedia, The Free Encyclopedia. URL: <http://en.wikipedia.org/w/index.php?title=Folksonomy&oldid=439642758> (besucht am 04.04.2012).
- Gradmann, Stefan (2005). „rdfs:frbr: Towards an Implementation Model for Library Catalogs Using Semantic Web Technology“. In: Cataloging & Classification Quarterly 39.3/4, S. 63–76. ISSN: 1544-4554. DOI: 10.1300/J104v39n03\_05.
- Graves, Mike, Adam Constanbaris und Dan Brickley (2007). „FOAF: Connecting People on the Semantic Web“. In: Cataloging & Classification Quarterly 43.3, S. 191–202. ISSN: 0163-9374. DOI: 10.1300/J104v43n03\_10.
- Greenall, Rurik Thomas (2011). „A Novel Method for Creating a Distributed, Collaborative Commenting Environment for Bibliographic Items“. In: code{4}lib (14). ISSN: 1940-5758. URL: <http://journal.code4lib.org/articles/5339> (besucht am 04.04.2012).
- Guide to Open Data Licensing. Version 1.0. Open Knowledge Foundation. URL: <http://www.opendefinition.org/guide/data/> (besucht am 13.04.2012).
- Hakala, Juha und Hartmut Walravens (2001). Using International Standard Book Numbers as Uniform Resource Names. RFC 3187. 11 S. URL: <http://tools.ietf.org/pdf/rfc3187.pdf> (besucht am 04.04.2012).
- Hatcher, Jordan S. (2010). „Open data and the law“. In: Nodalities (9), S. 5–6. ISSN: 1757-2592. URL: [http://www.talis.com/nodalities/pdf/nodalities\\_issue9.pdf](http://www.talis.com/nodalities/pdf/nodalities_issue9.pdf) (besucht am 04.04.2012).
- Hauer, Manfred und Reiner Diedrichs (2009). „Vom Katalog zur Bibliothek: Zwischenschritt und Zwischenstand“. In: Bibliotheksdienst 43.10, S. 1031–1038. ISSN: 0006-1972. URL: [http://www.zlb.de/aktivitaeten/bd\\_neu/heftinhalte2009/Erschliessung021009\\_BD.pdf](http://www.zlb.de/aktivitaeten/bd_neu/heftinhalte2009/Erschliessung021009_BD.pdf) (besucht am 04.04.2012).

- 
- Hauke, Petra, Aline Hötzeltdt und Jana Rumler (2010). „Kataloganreicherung?! – Vom Zugang zu Open-Access-Publikationen über Bibliothekskataloge am Beispiel hybrider Publikationen“. In: Bibliothek. Forschung und Praxis 34.3, S. 293–305. ISSN: 1865-7648. DOI: 10.1515/bfup.2010.045.
- Hauke, Petra, Jana Rumler und Aline Hötzeltdt (2010). „Open Access - nutzerfreundlich!? Nachweise paralleler Druck- und Online-Ausgaben in Bibliothekskatalogen und Dokumenten-Suchmaschinen“. In: ABI Technik 30.3, S. 174–181. ISSN: 0720-6763.
- Heath, Tom (2010). „Applying Licenses and Waivers to Linked Data“. In: Nodalities (9), S. 15–17. ISSN: 1757-2592. URL: [http://www.talis.com/nodalities/pdf/nodalities\\_issue9.pdf](http://www.talis.com/nodalities/pdf/nodalities_issue9.pdf) (besucht am 04.04.2012).
- Heath, Tom und Christian Bizer (2011). Linked Data. Evolving the Web into a Global Data Space. ISBN: 9781608454310. DOI: 10.2200/S00334ED1V01Y201102WBE001.
- Increase Sales & Lower Costs With Better Metadata (2010). A Step-By-Step ONIX Guide for Publishers. Last Updated June 2010; Prepared by BookNet Canada Staff and Canadian Bookshelf Staff. URL: [http://www.booknetcanada.ca/images/media/StC/bnc\\_onix\\_marketing\\_v1.pdf](http://www.booknetcanada.ca/images/media/StC/bnc_onix_marketing_v1.pdf) (besucht am 04.04.2012).
- International Federation of Library Associations and Institutions / Study Group on the Functional Requirements for Bibliographic Records ; Deutsche Nationalbibliothek <Leipzig; Frankfurt, Main>, Hrsg. (2009). Funktionale Anforderungen an bibliografische Datensätze. Abschlussbericht der IFLA Study Group on the Functional Requirements for Bibliographic Records. Geänderte und korrigierte Fassung, Stand: Februar 2009. Leipzig ; Frankfurt, M. ; Berlin: Dt. Nationalbibliothek. viii, 138 S. ISBN: 978-3-941113-05-3. URN: <http://nbn-resolving.de/urn:nbn:de:101-2009022600> (besucht am 04.04.2012).
- Jacobson, Kurt, Yves Raimond und Thomas Gängler (2010). The Similarity Ontology - MuSim. Working Draft. Hrsg. von Kurt Jacobson. Version V. 0.2, Date: 23:06:39 10/03/10 BST. URL: <http://kakapo.dcs.qmul.ac.uk/ontology/musim/0.2/musim.html> (besucht am 13.04.2012).

- 
- Jaffri, Afraz, Hugh Glaser und Ian Millard (2007). „URI Identity Management for Semantic Web Data Integration and Linkage“. In: 3rd International Workshop On Scalable Semantic Web Knowledge Base Systems (25.–30. Nov. 2007). Vilamoura, Algarve, Portugal. URL: <http://eprints.ecs.soton.ac.uk/14361/> (besucht am 04.04.2012).
- Jochum, Uwe (2007). Kleine Bibliotheksgeschichte. 3., verb. u. erw. Aufl. Ditzingen: Reclam. 280 S. ISBN: 978-3-15-017667-2.
- JSON. Bearbeitungsstand: 23. August 2011, 17:23 UTC. In: Wikipedia, The Free Encyclopedia. URL: <http://en.wikipedia.org/w/index.php?title=JSON&oldid=446347656> (besucht am 04.04.2012).
- JSONP. Bearbeitungsstand: 18. August 2011, 21:37 UTC. In: Wikipedia, The Free Encyclopedia. URL: <http://en.wikipedia.org/w/index.php?title=JSONP&oldid=445568624> (besucht am 04.04.2012).
- Kieft, Robert (2004). Collaborative Project to Enhance Library Catalog Browsing. URL: <http://www.loc.gov/standards/catenrich/catenrich-clir.html> (besucht am 20.02.2011).
- Kneifel, Fabienne (2009). Mit Web 2.0 zum Online-Katalog der nächsten Generation. B.I.T.online – Innovativ 23. Wiesbaden: Dinges & Frick. 160 S. ISBN: 978-3-934997-26-4. URL: [http://www.b-i-t-online.de/daten/BIT\\_Innovativ\\_23\\_Kneifel.pdf](http://www.b-i-t-online.de/daten/BIT_Innovativ_23_Kneifel.pdf) (besucht am 04.04.2012).
- Knull-Schlomann, Kristina (2009). „Der Mehrwert bibliografischer Datensätze“. In: Dialog mit Bibliotheken 21.2, S. 40–43. ISSN: 0936-1138. URL: [http://files.d-nb.de/pdf/dialog\\_09\\_2\\_vollt.pdf](http://files.d-nb.de/pdf/dialog_09_2_vollt.pdf) (besucht am 04.04.2012).
- Koreferenzen im Semantic Web (2011). Version V. 40, Apr 26, 2011 08:53. URL: <https://wiki1.hbz-nrw.de/display/SEM/Koreferenzen+im+Semantic+Web> (besucht am 04.04.2012).
- Korn, Naomi und Charles Oppenheim (Mai 2011). Licensing Open Data: A Practical Guide. URL: [http://discovery.ac.uk/files/pdf/Licensing\\_Open\\_Data\\_A\\_Practical\\_Guide.pdf](http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf) (besucht am 04.04.2012).

- 
- Kreutzer, Till (2011). Digitalisierung gemeinfreier Werke durch Bibliotheken. Ein Leitfaden von Dr. Till Kreutzer; hg. v. Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen. 51 S. URL: <http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/Digitalisierungsleitfaden.pdf> (besucht am 04.04.2012).
- Löhner, Sabina (2007). Kataloganreicherung in Hochschulbibliotheken. State of the Art Übersicht und Aussichten für die Schweiz. Churer Schriften zur Informationswissenschaft 15. 124 S. URL: <http://e-collection.ethbib.ethz.ch/eserv.php?pid=eth:29340&dsID=eth-29340-01.pdf> (besucht am 04.04.2012)..
- Library mashups (2009). exploring new ways to deliver library data. Engard, Nicole C, Hrsg. Medford, N.J.: Information Today, Inc. xvii, 334 S. ISBN: 9781573873727.
- Library of Congress Network Development and MARC Standards Office, Hrsg. (Jan. 2000). Record Structure: MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media. Version 2007-12-05. URL: <http://www.loc.gov/marc/specifications/specrecstruc.html> (besucht am 04.04.2012).
- Möller, Erik (2005). Freiheit mit Fallstricken. Die Creative Commons NC-Lizenzen und ihre Folgen. URL: <http://freedomdefined.org/Licenses/NC/De> (besucht am 26.08.2011).
- Manning, Christopher D., Prabhakar Raghavan und Hinrich Schütze (2009). Introduction to information retrieval. Reprinted. Literaturverz. S. 441 - 468. Cambridge [u.a.]: Cambridge Univ. Press. XXI, 482 S. ISBN: 978-0-521-86571-5. URL: <http://nlp.stanford.edu/IR-book/information-retrieval-book.html> (besucht am 04.04.2012).
- Markner, Reinhard (2. Mai 2005). „Kampfplatz Katalog. die Verzeichnisse der Bibliotheken werden fragwürdig „angereichert““. In: Süddeutsche Zeitung 100, S. 16. URL: <http://markner.free.fr/kampf.htm> (besucht am 04.04.2012).
- Mendes, Luiz H., Jennie Quinonez-Skinner und Danielle Skaggs (2009). „Subjecting the catalog to tagging“. In: Library Hi Tech 27.1, S. 30–41. ISSN: 0737-8831. DOI: 10.1108/07378830910942892.
- Mi, Jia und Cathy Weng (2008). „Revitalizing the Library OPAC. Interface, Searching, and Display Challenges“. In: Information Technology and Libraries 27.1 (March), S. 5–22. ISSN: 0730-9295. URL: <http://www.ala.org/lita/ital/files/27/1/mi.pdf> (besucht am 13.04.2012).



- 
- Morgan, Eric Lease (1. Juni 2011). Next-generation library catalogs, or ‘Are we there yet?’ Blog-Eintrag. URL: <http://infomotions.com/blog/2011/06/next-generation-library-catalogs-or-are-we-there-yet/> (besucht am 04.04.2012).
- Morris, Ruth C. (2001). „Online tables of contents for books: effect on usage“. In: Bulletin of the Medical Library Association 89.1, S. 29–36. ISSN: 0025-7338. URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC31701/> (besucht am 04.04.2012).
- „Named Graphs“ (2005). Carroll, Jeremy J. et al. In: Journal of Web Semantics 3.4, S. 247–267. ISSN: 1570-8268. DOI: 10.1016/j.websem.2005.09.001.
- O’Brien, Ann (1994). „Online Catalogs: Enhancements and Developments“. In: Annual Review of Information Science and Technology (ARIST) 29, S. 219–242. ISSN: 0066-4200.
- OCLC Online Computer Library Center, Inc. (2009). Online Catalogs: What Users and Librarians Want. Dublin, Ohio: OCLC. vi, 58 S. ISBN: 978-1-55653-411-9. URL: <http://www.oclc.org/reports/onlinecatalogs/default.htm> (besucht am 04.04.2012).
- O’Connor, Brian C. und Mary K. O’Connor (7. Sep. 1998). „Book Jacket as Access Mechanism. An Attribute Rich Resource for Functional Access to Academic Books“. In: First Monday 3.9. URL: <http://ojphi.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/616/537> (besucht am 04.04.2012).
- ONIX for Books (8. Apr.2009). Product Information Format; Specification. EDItEUR, Hrsg. Release 3.0. Version 2011-05-31. 154 S. URL: [http://www.editeur.org/files/ONIX%203/ONIX\\_for\\_Books\\_Release3-0\\_docs+codes\\_Issue\\_14.zip](http://www.editeur.org/files/ONIX%203/ONIX_for_Books_Release3-0_docs+codes_Issue_14.zip) (besucht am 04.04.2012).
- O’Steen, Ben (17. Okt. 2010). ”Bundling” instances of author names together without using owl:sameas. Blog-Eintrag. URL: <http://openbiblio.net/2010/11/17/bundling-instances-of-author-names-together-without-using-owlsameas/> (besucht am 04.04.2012).
- Page, Roderic D. M. (2010). „Enhanced display of scientific articles using extended metadata“. In: Journal of Web Semantics 8.2. Preprint. ISSN: 1570-8268. URL: <http://www.websemanticsjournal.org/index.php/jps/article/view/14> (besucht am 04.04.2012).

- 
- Pohl, Adrian (2010). Open Data im hbz-Verbund. 12 S. URL: [http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl\\_2010\\_open-data.pdf](http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_2010_open-data.pdf) (besucht am 04.04.2012).
- Pohl, Adrian (2011). Linked Data und die Bibliothekswelt. 22 S. URL: [http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl\\_2011\\_linked-data\\_ODOK.pdf](http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_2011_linked-data_ODOK.pdf) (besucht am 04.04.2012).
- Pohl, Bettina (2006). Abstracts und andere Inhaltsmitteilungen im Urheberrecht. Osnabrück, Univ., Diss. XXX, 239 S. eprint: <http://repositorium.uni-osnabrueck.de/handle/urn:nbn:de:gbv:700-2006112213> (besucht am 13.04.2012).
- Precision and recall. Bearbeitungsstand: 6. Juni 2011, 05:27 UTC. In: Wikipedia, The Free Encyclopedia. URL: [http://en.wikipedia.org/w/index.php?title=Precision\\_and\\_recall&oldid=432801351](http://en.wikipedia.org/w/index.php?title=Precision_and_recall&oldid=432801351) (besucht am 04.04.2012).
- Principle of least effort. Bearbeitungsstand: 26. Juni 2010, 08:06 UTC. In: Wikipedia, The Free Encyclopedia. URL: [http://en.wikipedia.org/w/index.php?title=Principle\\_of\\_least\\_effort&oldid=370220388](http://en.wikipedia.org/w/index.php?title=Principle_of_least_effort&oldid=370220388) (besucht am 04.04.2012).
- Ranganathan, Shiyali R. (1989). The Five Laws of Library Science. Bangalore (Indien): Sarada Ranganathan Endowment. 449 S. ISBN: 81-85273-08-1.
- RDF Primer (10. Feb. 2004). W3C Recommendation 10 February 2004. Manola, Frank und Eric Miller, Hrsg. Version 20040210. W3C. URL: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/> (besucht am 04.04.2012).
- Same origin policy. Bearbeitungsstand: 19. August 2011, 04:23 UTC. In: Wikipedia, The Free Encyclopedia. URL: [http://en.wikipedia.org/w/index.php?title=Same\\_origin\\_policy&oldid=445615584](http://en.wikipedia.org/w/index.php?title=Same_origin_policy&oldid=445615584) (besucht am 04.04.2012).
- Samulski, Roland (2007). „Die Rolle des Catalogue Enrichment im Rahmen der Entwicklungsperspektiven von Online-Katalogen als Dienstleistungsprodukte. Eine Brücke in die Zukunft oder eine Investition in ein überholtes Modell?“ Masterarbeit. FH Köln, Institut für Informationswissenschaft. 94 S.

- 
- Schöllhorn, Katharina (2009). „Normdaten in Onlinekatalogen – Bestandsaufnahme und Entwicklungschancen“. Bachelorarbeit. Hochschule der Medien Stuttgart. 97 S. URL: [http://opus.bsz-bw.de/hdms/volltexte/2009/679/pdf/schoellhorn\\_Bachelorarbeit\\_normdaten\\_onlinekatalogen.pdf](http://opus.bsz-bw.de/hdms/volltexte/2009/679/pdf/schoellhorn_Bachelorarbeit_normdaten_onlinekatalogen.pdf) (besucht am 04.04.2012).
- Seiffert, Florian (2001). „Eine Analyse der Verbunddaten des hbz“. In: ABI-Technik 21.2, S. 125–146. ISSN: 0720-6763.
- Semantisches Web. Bearbeitungsstand: 30. Juli 2011, 14:04 UTC. In: Wikipedia, Die freie Enzyklopädie. URL: [http://de.wikipedia.org/w/index.php?title=Semantisches\\_Web&oldid=91885476](http://de.wikipedia.org/w/index.php?title=Semantisches_Web&oldid=91885476) (besucht am 04.04.2012).
- Spalding, Tim (15. März 2007). thingISBN data in one file. URL: <http://www.librarything.com/blogs/thingology/2007/03/thingisbn-data-in-one-file/> (besucht am 04.04.2012).
- Spiteri, Lousie F. (2007). „The Structure and Form of Folksonomy Tags: The Road to the Public Library Catalog“. In: Information Technology and Libraries 26.3, S. 13–25. ISSN: 0730-9295. URL: <http://www.ala.org/lita/ital/files/26/3/spiteri.pdf> (besucht am 13.04.2012).
- Tag (metadata). Bearbeitungsstand: 5. Juni 2011, 15:49 UTC. In: Wikipedia, The Free Encyclopedia. URL: [http://en.wikipedia.org/w/index.php?title=Tag\\_\(metadata\)&oldid=432696380](http://en.wikipedia.org/w/index.php?title=Tag_(metadata)&oldid=432696380) (besucht am 04.04.2012).
- Tennant, Roy (2005). „Lipstick on a Pig“. In: Library Journal 130.7, S. 34. ISSN: 0363-0277. URL: <http://www.libraryjournal.com/article/CA516027.html> (besucht am 13.04.2012).
- Thomas, Marliese, Dana M. Caudle und Cecilia Schmitz (2010). „Trashy tags: problematic tags in LibraryThing“. In: New Library World 111.5/6, S. 223–235. ISSN: 0307-4803. DOI: 10.1108/03074801011044098.
- Van Orden, Richard (1990). „Content-enriched access to electronic information: Summaries of Selected Research“. In: Library Hi Tech 8.3, S. 27–32. ISSN: 0737-8831. DOI: 10.1108/eb047795.
- Voß, Jakob (25. Apr. 2008a). Bibliographic Hash Key. Version 2011-04-05, 11:46 Uhr. URL: [http://www.gbv.de/wikis/cls/index.php?title=Bibliographic\\_Hash\\_Key&oldid=13230](http://www.gbv.de/wikis/cls/index.php?title=Bibliographic_Hash_Key&oldid=13230) (besucht am 13.04.2012).

- 
- Voß, Jakob (2008b). „SeeAlso: A Simple Linkserver Protocol“. In: Ariadne 57. ISSN: 1361-3200. URL: <http://www.ariadne.ac.uk/issue57/voss/> (besucht am 13.04.2012).
- Voß, Jakob (12. Feb. 2009). SeeAlso Simple Specification. URL: [http://www.gbv.de/wikis/cls/SeeAlso\\_Simple\\_Specification](http://www.gbv.de/wikis/cls/SeeAlso_Simple_Specification) (besucht am 04.04.2012).
- Wendelin, Karla Hawkins und R. Ann Zinck (1983). „How Students Make Book Choices“. In: Reading Horizons 23.2, S. 84–88. ISSN: 0034-0502.
- „When owl:sameAs isn’t the Same (2010): An Analysis of Identity in Linked Data“. Halpin, Harry et al. In: 9th International Semantic Web Conference (ISWC2010) (7.–11. Nov. 2010). Shanghai, China. URL: <http://data.semanticweb.org/conference/iswc/2010/paper/261> (besucht am 04.04.2012).
- Why Open Up Bibliographic Data. Open Knowledge Foundation. URL: <http://openbiblio.net/why/> (besucht am 04.04.2012).
- Wiesenmüller, Heidrun (2008). „Zehn Jahre Functional Requirements for Bibliographic Records”(FRBR). Vision, Theorie und praktische Anwendung“. In: Bibliothek. Forschung und Praxis 32.3, S. 348–359. ISSN: 1865-7648. DOI: 10.1515/bfup.2008.047.
- Yee, Raymond (2008). Pro Web 2.0 Mashups. Remixing Data and Web Services. Berkeley, CA : Apress. xxxiii, 603 S. ISBN: 978-1-59059-858-0 (pbk.), 978-1-4302-0286-8 (electron.) URL: <http://mashupguide.net/1.0/html/> (besucht am 04.04.2012).
- Yu, Holly und Margo Young (2004). „The Impact of Web Search Engines on Subject Searching in OPAC“. In: Information Technology and Libraries 23.4 (December), S. 168–180. ISSN: 0730-9295. URL: <http://www.ala.org/lita/ital/files/23/4/234yu.pdf> (besucht am 13.04.2012).

---

## 14 CODE-LISTINGS

In den Listings 4 und 6 ist `${query}` jeweils durch den Suchbegriff (Identifikator) zu ersetzen.

```
01 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
02 @prefix coref: <http://www.rkbexplorer.com/ontologies/coref#> .
03
04 <http://caesar.example.org/bundle/1>
05 ## Kanonische URI - Vorzugsbezeichnung
06 # coref:canon <...> ;
07 ## Zeitstempel letzte Aktualisierung
08 # coref:lastUpdated "2011-08-29 17:51:10" ;
09   coref:duplicate
10     <urn:nbn:de:eki:HBZHT016905880> ,
11     <urn:isbn:978-0-12-385965-5> ,
12     <urn:isbn:0-12-385965-4> ,
13     <http://193.30.112.134/F?func=find-c&ccl_term=IDN%3DHT016905880> ,
14     <urn:oclc:734070871> ,
15     <urn:oclc:712780761> ,
16     <http://lccn.loc.gov/2011010645> ,
17     <http://openlibrary.org/books/OL24839081M> ;
18   a coref:Bundle .
```

**Listing 3** coref-Datensatz (in Turtle-Syntax)

```
01 PREFIX coref: <http://www.rkbexplorer.com/ontologies/coref#>
02
03 SELECT ?id
04 WHERE {
05   ?bundle coref:duplicate ${query} .
06   ?bundle coref:duplicate ?id .
07 }
```

**Listing 4** SPARQL-Abfrage der zu einem coref-Bundle gehörenden Identifikatoren anhand eines bekannten Identifikators

```

01 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
02 @prefix dc: <http://purl.org/dc/elements/1.1/> .
03 @prefix caesar: <http://caesar.example.org/vocab/enrichments#> .
04 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
05 @prefix bibo: <http://purl.org/ontology/bibo/> .
06 @prefix scom:
07   <http://sites.wiwiiss.fu-berlin.de/suhl/bizer/bookmashup/simpleCommerceVocab01.rdf#> .
08
09 <http://loc.gov/> {
10   <http://lccn.loc.gov/2011010645>
11     dc:subject
12       "Semantic Web", "Web site development", "Metadata" ;
13     caesar:ToC
14       "What is the Semantic Web? -- Semantic modeling -- RDFS -- the basis of the
15       Semantic Web -- Semantic Web application architecture -- Querying the Semantic
16       Web -- SPARQL -- RDF and inferencing -- RDF schema -- RDFS-Plus -- Using RDFS-
17       Plus in the wild -- SKOS -- managing vocabularies with RDFS-Plus -- Basic OWL --
18       Counting and sets in OWL -- Ontologies on the Web -- putting it all together --
19       Good and bad modeling practices -- Expert modeling in OWL -- Conclusions."
20 }
21
22 <http://www.amazon.de/> {
23   <urn:isbn:978-0-12-385965-5>
24     foaf:depiction
25       <http://ecx.images-amazon.com/images/I/51CI0qnpeXL._SL160_.jpg> ;
26     scom:hasOffer
27       <http://www.amazon.de/exec/obidos/ASIN/0123859654> .
28
29   <urn:isbn:0-12-385965-4>
30     foaf:depiction
31       <http://ecx.images-amazon.com/images/I/51CI0qnpeXL._SL160_.jpg> ;
32     scom:hasOffer
33       <http://www.amazon.de/exec/obidos/ASIN/0123859654> .
34 }
35
36 <http://books.google.com/> {
37   <urn:isbn:978-0-12-385965-5>
38     caesar:limitedFreePreview
39       <http://books.google.com/books?id=_qGKP01B1DgC> ;
40     bibo:abstract
41       ""Semantic Web models and technologies provide information in machine-
42       readable languages that enable computers to access the Web more int..."" ;
43     foaf:depiction
44       <http://books.google.
45       com/books?id=_qGKP01B1DgC&printsec=frontcover&hl=de&source=gbs_ge_summary_r&cad=0> .
46 }
47
48 <http://caesar.example.org/enrichments/> {
49   <http://caesar.example.org/bundle/1>
50     caesar:tag
51       "ontology", "rdf" .
52 }

```

Listing 5 Anreicherungen in Benannten Graphen (in TriG-Syntax)

```

01 SELECT ?src ?type ?res
02 WHERE {
03   GRAPH ?src { ${query} ?type ?res . }
04 }

```

Listing 6 SPARQL-Abfrage der Anreicherungen zu einem Identifikator

```

01 #!/usr/bin/perl
02 use strict;
03 use warnings;
04 use RDF::Trine;
05 use RDF::Trine::Parser;
06 use RDF::Query;
07 my $model = RDF::Trine::Model->temporary_model();
08
09 my $ttl_parser = RDF::Trine::Parser::Turtle->new();
10 $ttl_parser->parse_file_into_model( "http://caeser.example.org/", "koreferenzen.ttl", ➤
11 $model );
12 printf ("%3d Statements im Speicher\n", $model->count_statements() );
13
14 my $trig_parser = RDF::Trine::Parser::TriG->new();
15 $trig_parser->parse_file_into_model( undef, "anreicherungen.trig", $model );
16 printf ("%3d Statements im Speicher\n", $model->count_statements() );
17
18 my $qid = "urn:nbn:de:eki:HBZHT016905880";
19 my $sel_ids = <<"GETID";
20 PREFIX coref: <http://www.rkbexplorer.com/ontologies/coref#>
21 SELECT ?id
22 WHERE {
23     ?bundle coref:duplicate <$qid> .
24     ?bundle coref:duplicate ?id .
25 }
26 GETID
27
28 my $query = RDF::Query->new($sel_ids);
29 my $iter = $query->execute($model);
30 while ( my $coref = $iter->() ) {
31     my $e_id = $coref->{'id'};
32     printf ( "Koreferente URI: %s\n", $e_id );
33     my $sel_enrich = <<"GETENRICH";
34 PREFIX coref: <http://www.rkbexplorer.com/ontologies/coref#>
35 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
36 SELECT ?src ?type ?res
37 WHERE {
38     GRAPH ?src { $e_id ?type ?res . }
39 }
40 GETENRICH
41
42 my $query = RDF::Query->new($sel_enrich);
43 my $iter = $query->execute($model);
44 while ( my $e = $iter->() ) {
45     printf (<<"END", $e->{'type'}, $e->{'res'}, $e->{'src'}, );
46     Anreicherung gefunden!
47     Typ      : %s
48     Wert     : %s
49     Ursprung : %s
50 END
51 }
52 }

```

**Listing 7** Perl-Skript zum Laden der in den vorstehenden Listings dargestellten Daten und zur nachfolgenden Abfrage per SPARQL